

PCT

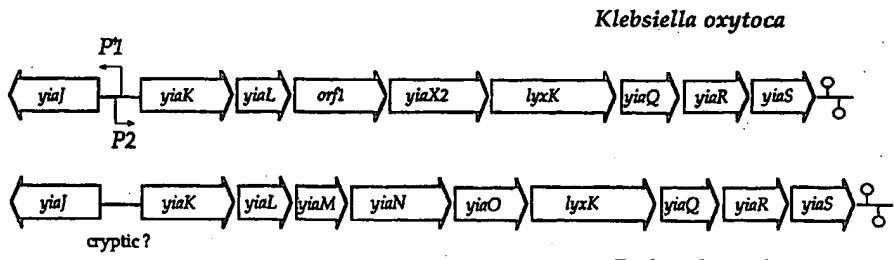
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68, C12P 21/06, C12N 15/00, C07H 21/02, 21/04, A61K 39/108		A1	(11) International Publication Number: WO 00/22170 (43) International Publication Date: 20 April 2000 (20.04.00)
(21) International Application Number: PCT/US99/23862 (22) International Filing Date: 12 October 1999 (12.10.99)		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 09/172,952 14 October 1998 (14.10.98) US			
(71) Applicant: MICROGENOMICS, INC. [US/US]; 11211 Sorrento Valley Road, San Diego, CA 92121 (US). (72) Inventors: HOCH, James; 1054 Havenhurst Drive, La Jolla, CA 92037 (US). DARTOIS, Veronique; Apartment 233, 3957 Nobel Drive, San Diego, CA 92122 (US). (74) Agent: AMES, Wesley, B.; Lyon & Lyon LLP, Suite 4700, 633 West Fifth Street, Los Angeles, CA 90071-2066 (US).			

(54) Title: METABOLIC SELECTION METHODS



(57) Abstract

The present invention relates in part to methods for screening for novel enzymatic pathways in environmental samples using metabolic selection strategies, and the isolation of the genes and proteins that make up these pathways, and the figure compares a physical map of *Klebsiella oxytoca* yiaK-S operon with the *E. coli* yaiK-S operon.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

DESCRIPTIONMetabolic Selection MethodsField Of The Invention

The present invention relates to methods for screening for enzymatic pathways, and the isolation of the genes and proteins that make up these pathways.

Background Of The Invention

The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to be, or to describe, prior art to the invention.

Biological synthesis of compounds is frequently more cost effective and more productive than chemical synthesis, which can have low yields, require expensive and toxic reagents, and require lengthy purifications. In contrast, biological synthesis using known pathways can be rapid, with high yields. However, the identification of new biological pathways for syntheses of interest is difficult and time consuming.

Currently, the biochemical screening of isolates is a major means by which people find new pathways for the production of chemicals, antibacterials, and other anti-infectives. However, screening is inherently several orders of magnitude slower than selection and requires that the organism be cultured in the laboratory. Since at least 99% of the microbes in the environment do not grow on laboratory media, less than 1% can be tested using a biochemical screen. Thus, biological pathways in 99% of organisms will never be found by classical biochemical screening technologies.

Summary Of The Invention

The metabolic selection strategy of this invention is designed to find an enzymatic pathway for the conversion of any

source compound to any target compound. Conservatively, this technique allows at least a million-fold increase in the discovery rate over classical biochemical screening approaches, and allows testing of the 99% of the environmental microbes that 5 are currently unable to be cultured in the laboratory.

A biocatalytic or metabolic pathway consists of a series of protein catalysts (enzymes) which catalyze the conversion of a starting material to the final product. A general process to identify the metabolic pathway from a source compound to a target 10 compound involves the creation/ identification of an easily genetically-manipulatable organism containing an inducible signal, which is activated when a target compound is metabolized. This is followed by the screening of nucleic acid in this organism to identify genes which metabolize the source compound 15 to the target compound.

An example of a selection strategy which can be used to identify the metabolic pathway from a source compound to a target compound is diagrammed in Figure 11. As a first step, microbial isolates are selected that are capable of metabolizing a target 20 compound "T", but not a source compound "S", to an essential factor. Essential factors can include elements like carbon, sulfur, phosphorous, and nitrogen, or other essential nutrients, e.g. some amino acids, fatty acids, and carbohydrates. In a second step, the pathway responsible for the catabolism of 25 compound "T" is identified and made conditional. That is, the gene(s) for the pathway is cloned and placed under control of an inducible promoter such that growth on the target compound is turned "ON" only when the inducer is present. This engineered strain is referred to as the "tester strain". The third part of 30 the strategy is the transfer of foreign DNA from environmental sources into the tester strain, followed by selection for growth on the source compound "S" in the presence of inducer. Such positive clones either are capable of metabolizing compound "S" in the absence of inducer, in which case utilization of "S" does 35 not require prior conversion to compound "T" (Figure 11; pathway

I), or alternatively metabolize compound "S" only when "T" catabolism is "ON", suggesting that utilization of "S" proceeds via compound "T" to intermediary metabolism (Figure 11; pathway II). These latter clones are further analyzed and the 5 biocatalysts for the conversion of "S" to "T" are characterized.

A specific embodiment of the metabolic selection strategy is shown in Figure 12, where "S" is 2-keto-L-gulonate (2-KLG), and "T" is ascorbic acid (AsA) which can be metabolized to carbon and energy.

10 Thus, in a first aspect, the invention features a method of screening for one or more nucleic acid sequences which express a product or products that convert a source compound into a target compound. The method comprises contacting a cell with one or more test nucleic acid sequences, where the cell expresses one or 15 more genes encoding one or more proteins which, in the presence of the target compound, provide a detectable signal. The detectable signal indicates the presence of the desired nucleic acid sequence or sequences.

20 The term "screening" as used herein refers to methods for identifying a nucleic acid sequence of interest. Preferably, the method permits the identification of a nucleic acid sequence of interest among one or more sequences, more preferably among hundreds (100, 200,...900), most preferably among thousands (1,000, 2,000,...etc.) or more. The sequences to be screened can 25 be isolated from one or more organisms. Preferably, the sequences are isolated from hundreds of organisms, more preferably from thousands or more organisms. The term "screening" may include both classical screening, whereby expression of the nucleic acid results in a phenotype that can be 30 identified (for example by having a colony with the nucleic acid of interest change color, fluoresce, or luminesce), and may also include classical selection, where typically the phenotype to be identified is growth on selective media. By "selective" is meant media on which the host strain will not grow or grows poorly, but 35 that strains with the nucleic acid of interest will grow in a

manner which can be readily distinguished from host strain growth by methods well-known in the art.

The term "nucleic acid" as used herein refers to either deoxyribonucleic acid or ribonucleic acid that may be isolated, enriched, or purified from natural sources or synthesized recombinantly. These methods are well-known in the art and specific examples are also given herein. Preferably, a "nucleic acid" to be identified in the screening method comprises a nucleic acid encoding a metabolic pathway that is not normally found in the cell. Thus, preferably, the pathway has not simply been inactivated through a mutation and the relevant genes are now being identified through complementation. Rather the nucleic acid being identified does not normally exist in the cell in which it is being screened for. Typically, the screening is cross strains, more typically, cross-species, and even more preferably, cross-genera or with further remoteness.

By "isolated, purified, or enriched" in reference to nucleic acid is meant a polymer of 6 (preferably 21, more preferably 39, most preferably 75) or more nucleotides conjugated to each other, including DNA and RNA that is isolated from a natural source or that is synthesized. In certain embodiments of the invention, longer nucleic acids are preferred, for example those of 300, 600, 900 or more nucleotides and/or those having at least 50%, 60%, 75%, 90%, 95% or 99% identity to the sequence shown in SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, or SEQ ID NO:19.

The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular (i.e., chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide chain present, but that it is essentially free (about 90-95% pure

at least) of non-nucleotide material naturally associated with it, and thus is distinguished from isolated chromosomes.

By the use of the term "enriched" in reference to nucleic acid is meant that the specific DNA or RNA sequence constitutes 5 a significantly higher fraction (2-5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a 10 preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two. However, it should be noted that "enriched" does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term 15 "significant" is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other nucleic acids of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no DNA or RNA from other 20 sources. The other source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other 25 species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference 30 to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, e.g., in terms of mg/mL). Individual 35 clones isolated from a cDNA library may be purified to

electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6 -fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

The term "expresses a product" as used herein refers to the production of proteins from a nucleic acid vector containing genes within a cell. The nucleic acid vector is transfected into cells using well known techniques in the art as described herein. The "product" may, or may not, be naturally present in the cell.

The term "nucleic acid vector" relates to a single- or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a desired product can be inserted into a vector by cutting the vector with restriction enzymes and ligating the pieces together, depending on the availability of useful restriction sites. However, there are many methods well-known in the art for the insertion of nucleic acid sequences into vectors.

The term "transfecting" as used herein includes a number of methods to insert a nucleic acid vector or other nucleic acid molecules into a cellular organism. These methods involve a variety of techniques, such as treating the cells with high 5 concentrations of salt, an electric field, detergent, or DMSO to render the outer membrane or wall of the cells permeable to nucleic acid molecules of interest or use of various viral transduction strategies.

The term "converts" as used herein refers to changing one 10 compound into another compound, preferably enzymatically. The "source compound" refers to the compound to be converted to the "target compound." The "target compound" includes not only the compound that is metabolized to form a detectable signal, but can also include intermediates along the path to a detectable signal. 15 This is particularly preferred if the target compound is a surrogate target. By "surrogate target compound" is meant a target that is used because the preferable target cannot be used for any of several potential reasons (e.g. if it doesn't cross membranes, has a short half-life, easily broken down, etc.). The 20 "target compound" also includes interconvertible compounds. By "interconvertible" is meant that a pathway exists in the tester strain to convert the compound to the target compound.

The term "contacting" as used herein refers to mixing a solution comprising the test nucleic acid with a liquid medium 25 bathing the cells of the methods. The solution comprising the nucleic acid may also comprise other components, such as dimethyl sulfoxide (DMSO), which facilitates the uptake of the test nucleic acid into the cells of the methods. This may also be done by other methods well-known in the art including, but not 30 limited to, transfection or transformation techniques. The solution comprising the test nucleic acid may be added to the medium bathing the cells by utilizing a delivery apparatus, such as a pipet-based device or syringe-based device.

The term "cell" as used herein includes the typical 35 definition of a cell, and is further specifically intended to

include "cell-free" systems comprising the cellular machinery necessary to express the nucleic acid of the invention. By "cellular machinery" is meant the cellular components present in cell-free transcription and/or translation systems. Such systems 5 are well-known in the art. In particular, the "cell" lacks the ability to convert a source compound into a target compound, prior to the addition of test nucleic acid sequences. The term "lacks the ability" also includes cells in which the activity may be present but is at too low a level to provide a detectable 10 signal, or is low enough that an additional activity is detectably different. By "detectably different" is meant able to be measured over the background level (e.g. the level of the signal endogenously present in the "cell" and in the equipment used to measure the signal) by an amount greater than the level 15 of error present in the method of measuring.

The term "detectable signal" as used herein refers to a method of identification of the nucleic acids of interest e.g. by color, fluorescence, luminescence or growth.

In preferred embodiments of the method for screening nucleic 20 acid that converts a source compound into a target compound, the one or more nucleic acid sequences encodes a metabolic pathway not normally present in said cell. A "metabolic pathway" consists of a series of protein catalysts (enzymes) which catalyze the conversion of a starting material to a product. And 25 further, by "metabolic pathway" is meant the enzymes, and genes that encode them, that metabolize a source compound to a target compound.

In other preferred embodiments, the nucleic acid is selected 30 from the group consisting of mutagenized DNA, environmental DNA, combinatorial libraries, and recombinant DNA. Preferably, the environmental DNA is selected from the group consisting of mud, soil, sewage, flood control channels, sand, and water. Preferably the mutagenized DNA is the result of enzyme 35 mutagenesis where the mutagenesis is selected from the group consisting of random, chemical, PCR-based, and directed

mutagenesis. The directed mutagenesis is to include, for example, DNA shuffling. Preferably the enzymes to be mutagenized in this way are selected from the group consisting of lactonases, esterhydrolases, and reductases.

5 The term "environmental" as used herein refers to nucleic acids extracted from the environment, e.g. from mud, soil, or water. By "extracted" is meant isolated, enriched, or purified as defined above. The environmental sample can be directly extracted without prior laboratory culture, or can be pre-
10 cultured, for example, in the presence of a growth selective agent. Methods are known in the art and examples are described herein.

In still other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color. Methods for detecting these signals are well-known in the art. Preferably, the detectable signal is growth, and the target compound provides an element or factor required for growth.
20 Preferably the target compound is selected from the group consisting of ascorbate and 2-keto-L-gulonate (2-KLG), most preferably ascorbate. Preferably the element is selected from the group consisting of carbon, nitrogen, sulfur, and phosphorous. Most preferably, the element is carbon.
25 Alternatively, the essential factor is another essential nutrient. By "required for growth" is meant that the organism does not grow detectably in the absence of the element. By "provides an element" is meant that the compound can be metabolized by the organism, and that the result of this metabolism is the element in some form, e.g. carbon or carbon dioxide.

In other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the source compound is selected from the group consisting of 2-keto-L-gulonate (2-KLG), 2,5-deoxy-keto-gulonate
35

(2,5-DKG), L-idonate (L-IA), L-gulonate (L-GuA), and glucose, and most preferably 2-KLG.

In still other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the cell naturally expresses the one or more genes encoding one or more proteins that in the presence of the target compound provide a detectable signal. Alternatively, the cell can be genetically manipulated to express the one or more genes encoding one or more proteins that in the presence of the target compound provide a detectable signal. In both cases, the one or more proteins are preferably Yia operon-related polypeptides. The one or more genes are preferably under the control of an inducible promoter. The inducible promoter preferably comprises the trp-lac hybrid promoter, the lacO operator, and the lacI^q repressor.

By "naturally expresses" is meant that the genes encoding the proteins are present in the cell in its natural state, e.g. in nature, prior to culture in the laboratory. The genes may or may not be expressed in the natural state, or may or may not be expressed constitutively or inducibly. By "genetically manipulated to express" is meant the transfection of the desired genes into the cell by methods well-known in the art, examples of which are described herein.

The term "promoter" as used herein, refers to nucleic acid sequence needed for gene sequence expression. Promoter regions vary from organism to organism, but are well known to persons skilled in the art for different organisms. For example, in prokaryotes, the promoter region contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, ribosome binding site, start codon, and the like. By "inducible promoter" is meant a promoter which is only "on" in

the presence of an inducer. The "inducer" is typically a small molecule. Inducible promoters and inducers are well-known in the art and examples are given herein.

The term "Yia operon-related polypeptides" as used herein refers to polypeptides comprising 12 (preferably 15, more preferably 20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; 31 (preferably 35, more preferably 40, most preferably 50) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; 5 (preferably 10, more preferably 15, most preferably 25) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; 17 (preferably 20, more preferably 25, most preferably 35) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; 11 (preferably 15, more preferably 20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16; or a functional derivative thereof as described herein. In certain aspects, polypeptides of 100, 200, 300 or more amino acids are preferred. The Yia operon-related polypeptide can be encoded by its corresponding full-length nucleic acid sequence or any portion of its corresponding full-length nucleic acid sequence, so long as a functional activity of the polypeptide is retained (see, Examples section). It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide which retains the functionality of the original. In both cases, all permutations are within the embodiments of the invention.

The amino acid sequence of the Yia operon-related polypeptide will be substantially similar to the sequence shown in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID

NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18, or fragments thereof. A sequence that is substantially similar to the sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18 will preferably have at least 90% identity (more preferably at least 95% and most preferably 98-100%) to the sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18 using a Smith-Waterman protein-protein search.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues by the total number of residues and gaps and multiplying the product by 100. "Gaps" are spaces in an alignment that are the result of additions or deletions of amino acids. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity. For example, the computer algorithm BLAST is preferably used to search for homologous sequences in a database, and CLUSTAL is used to perform alignments. Identity and similarity determinations can be made using a Smith-Waterman protein-protein search, for example.

In still other preferred embodiments of the method for screening nucleic acid that converts a source compound into a target compound, the cell grows on ascorbate and does not grow on 2-KLG. Alternatively, the cell may grow on 2-KLG and not grow on 2,5-DKG. Preferably the cells are bacteria. Most preferably, the cell selective for ascorbate is *Klebsiella oxytoca*. By "grows on" is meant that the cell can utilize the compound (e.g. ascorbate or 2-KLG) as a source of carbon in the minimal essential media. However, the cell is unable to grow in the minimal essential media in the absence of the provided carbon

source. Thus, this provides a selective tool for the identification of the nucleic acid encoding the polypeptides of interest.

A second aspect of the invention features an isolated, enriched, or purified nucleic acid molecule encoding one or more 5 Yia operon-related polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

In preferred embodiments, the isolated, enriched, or purified nucleic acid molecule encoding one or more Yia operon-related polypeptides comprises a nucleotide sequence that: (a) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18; (b) is the complement of the nucleotide sequence of 10 (a); and (c) hybridizes under highly stringent conditions to the nucleotide molecule of (a) and encodes a naturally occurring 15 polypeptide.

In another preferred embodiment, the invention features an isolated, enriched, or purified nucleic acid molecule, wherein 20 said nucleic acid molecule comprises the nucleotide sequence set forth in SEQ ID NO:19. The nucleic acid molecule comprises: (a) one or more nucleotide sequences that are set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; (b) the 25 complement of the nucleotide sequence of (a); (c) nucleic acid that hybridizes under stringent conditions to the nucleotide molecule of (a); (d) the full length sequence of SEQ ID NO:19, except that it lacks one or more of the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, 30 SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; or (e) is the complement of the nucleotide sequence of (d).

The term "complement" refers to two nucleotides that can form multiple thermodynamically favorable interactions with one another. For example, adenine is complementary to thymine as 35 they can form two hydrogen bonds. Similarly, guanine and

cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if the nucleotides of the first sequence are complementary to the nucleotides of the second sequence. The 5 percent of complementarity (i.e. how many nucleotides from one strand form multiple thermodynamically favorable interactions with the other strand compared with the total number of nucleotides present in the sequence) indicates the extent of complementarity of two sequences.

10 Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. These conditions are well-known to those skilled in the art. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such 15 conditions prevent hybridization of nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides.

By "stringent hybridization conditions" is meant hybridization conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH₂PO₄, pH 6.8, 20 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhart's solution at 42 °C overnight; washing with 2X SSC, 0.1% SDS at 45 °C; and washing with 0.2X SSC, 0.1% SDS at 45 °C.

In other preferred embodiments the isolated, enriched, or purified nucleic acid molecule encoding one or more Yia operon-related polypeptides further comprises a vector or promoter effective to initiate transcription in a host cell. Preferably, the vector or promoter comprises the trp-lac hybrid promoter, the lacO operator, and the lacI^r repressor gene. In still other preferred embodiments, the nucleic acid molecule is isolated, 25 enriched, or purified from a bacteria, preferably *Klebsiella oxytoca*.

The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant nucleic acid may contain a sequence set forth in SEQ ID NO:1, SEQ ID 35 NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID

NO:7, SEQ ID NO:8, or SEQ ID NO:9, or a functional derivative thereof, and a vector or a promoter effective to initiate transcription in a host cell. The recombinant nucleic acid can alternatively contain a transcriptional initiation region 5 functional in a cell, a sequence complementary to an RNA sequence encoding one or more Yia operon-related polypeptides and a transcriptional termination region functional in a cell.

In preferred embodiments, the isolated, enriched, purified, recombinant, or recombinant in a cell, nucleic acid comprises, 10 consists essentially of, or consists of the full-length nucleic acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, encodes the full-length amino acid sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID 15 NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18, a functional derivative thereof, or at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. The Yia operon- 20 related polypeptides comprise, consist essentially of, or consist of at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. The nucleic acid may be isolated from a natural 25 source by cDNA cloning or by subtractive hybridization. The natural source may be prokaryotic, eukaryotic, or protozoal, preferably bacterial, from the environment, and the nucleic acid may be synthesized by the triester method or by using an automated DNA synthesizer. In other preferred embodiments, the 30 nucleic acid molecule is isolated, enriched, or purified from a bacteria, preferably *Klebsiella oxytoca*.

In yet other preferred embodiments, the nucleic acid is a conserved or unique region, for example those useful for: the design of hybridization probes to facilitate identification and 35 cloning of additional polypeptides, the design of PCR probes to

facilitate cloning of additional polypeptides, obtaining antibodies to polypeptide regions, and designing antisense oligonucleotides.

By "conserved nucleic acid regions", are meant regions present on two or more nucleic acids encoding a Yia operon-related polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions are provided in Abe, et al. (J. Biol. Chem. 19:13361-13368, 1992), hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables. Preferably, conserved regions differ by no more than 5 out of 20 nucleotides.

By "unique nucleic acid region" is meant a sequence present in a nucleic acid coding for a Yia operon-related polypeptide that is not present in a sequence coding for any other naturally occurring polypeptide. Such regions preferably encode 12 (preferably 15, more preferably 20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; 30 (preferably 35, more preferably 40, most preferably 50) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; 5 (preferably 10, more preferably 15, most preferably 25) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; 17 (preferably 20, more preferably 25, most preferably 35) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; 11 (preferably 15, more preferably 20, most preferably 30) or more contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16. In particular, a unique nucleic acid region is preferably of bacterial origin.

A third aspect of the invention features a nucleic acid probe for the detection of nucleic acid encoding one or more Yia operon-related polypeptides, selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS, in

a sample. Preferably, the nucleic acid probe encodes a polypeptide that is a fragment of the protein encoded by the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, 5 SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. The nucleic acid probe contains a nucleotide base sequence that will hybridize to the full-length sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, or a functional derivative thereof.

10 Hybridization is preferably under stringent conditions.

In preferred embodiments, the nucleic acid probe hybridizes to nucleic acid encoding at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; at least 30, 75, 15 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; at least 5, 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; at least 20 17, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; at least 11, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:16, 25 or a functional derivative thereof.

Methods for using the probes include detecting the presence or amount of Yia operon-related RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to Yia operon-related RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a Yia operon-related polypeptide may be used in the identification of the sequence of the nucleic acid detected (Nelson et al., in Non-isotopic DNA Probe Techniques, Academic Press, San Diego, Kricka, ed., p. 275, 1992, hereby incorporated

by reference herein in its entirety, including any drawings, figures, or tables). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

5 A fourth aspect of the invention features a recombinant cell comprising a nucleic acid molecule encoding one or more Yia operon-related polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS. In such cells, the nucleic acid may be under the control of the 10 genomic regulatory elements, or, preferably, may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" is meant a promoter that is not normally coupled *in vivo* transcriptionally to the coding sequence for the Yia operon-related polypeptides.

15 In preferred embodiments, the recombinant cell comprises nucleic acid encoding a polypeptide that is a fragment of the protein encoded by the amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. By 20 "fragment," is meant an amino acid sequence present in a Yia operon polypeptide. Preferably, such a sequence comprises at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; at least 30, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; at least 5, 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; at least 17, 32, 75, 90, 25 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; at least 11, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the 30 full-length amino acid sequence of SEQ ID NO:16.

Alternatively, the recombinant cell comprises the nucleic acid sequence set forth in SEQ ID NO:19, or comprises: (a) one or more nucleotide sequences that are set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ 5 ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; (b) the complement of the nucleotide sequence of (a); (c) nucleic acid that hybridizes under stringent conditions to the nucleotide molecule of (a); (d) the full length sequence of SEQ ID NO:19, except that it lacks one or more of the sequences set forth in SEQ ID NO:1, SEQ ID 10 NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; and (e) is the complement of the nucleotide sequence of (d). Preferably, the recombinant cell further comprises a vector or promoter effective to initiate transcription of the above-identified nucleic acid in the cell. 15 Preferably, the vector or promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene. Preferably, the recombinant cell is a bacteria, more preferably *Klebsiella oxytoca*.

Other preferred embodiments of this aspect of the invention 20 include a recombinant cell useful for screening for one or more nucleic acid sequences that express one or more products that convert a source compound into a target compound, where the cell expresses one or more genes, comprising an inducible promoter, and where the one or more genes encodes one or more proteins that 25 in the presence of the target compound and an inducer provide a detectable signal, where the detectable signal indicates the presence of the one or more nucleic acid sequences. Preferably, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color, and most 30 preferably is growth.

In preferred embodiments, of the recombinant cell useful for screening, the one or more nucleic acid sequences encodes a metabolic pathway not normally present in said cell. In other preferred embodiments, the nucleic acid is selected from the 35 group consisting of mutagenized DNA, environmental DNA,

combinatorial libraries, and recombinant DNA. Preferably, the environmental DNA is selected from the group consisting of mud, soil, sewage, flood control channels, sand, and water. Preferably the mutagenized DNA is the result of enzyme 5 mutagenesis where the mutagenesis is selected from the group consisting of random, chemical, PCR-based, and directed mutagenesis. The directed mutagenesis is to include, for example, DNA shuffling. Preferably the enzymes to be mutagenized in this way are selected from the group consisting of lactonases, 10 esterhydrolases, and reductases.

Additionally in this preferred embodiment, the cell preferably requires the presence of the target compound and the inducer for growth. Preferably, the target compound is selected from the group consisting of ascorbate and 2-KLG. In addition, 15 the one or more genes are preferably under the control of an inducible promoter, preferably comprising the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene. Preferably, the one or more proteins encoded by the one or more genes are one or more Yia operon-related polypeptides. 20 Preferably, the cell naturally expresses the one or more genes, or has been genetically manipulated to express the one or more genes. Preferably, the cell is a bacteria, most preferably *Klebsiella oxytoca*.

A fifth aspect of the invention features one or more 25 isolated, enriched, or purified Yia operon-related polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

By "isolated" in reference to a polypeptide is meant a polymer of 6 (preferably 12, more preferably 18, most preferably 30 25, 32, 40, or 50) or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are synthesized. In certain aspects longer polypeptides are preferred, such as those with 100, 200, 300, 400, or more contiguous amino acids of the sequence set forth in SEQ ID NO:10,

SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17 or SEQ ID NO:18.

The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or
5 separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid
10 chain present, but that it is essentially free (about 90-95% pure at least) of no-amino acid-based material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide is meant that the specific amino acid sequence
15 constitutes a significantly higher fraction (2-5 fold) of the total amino acid sequences present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acid
20 sequences present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been
25 significantly increased. The term significant here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acid sequences of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also
30 does not imply that there is no amino acid sequence from other sources. The other source of amino acid sequences may, for example, comprise amino acid sequence encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which man has intervened
35 to increase the proportion of the desired amino acid sequence.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication 5 that the sequence is relatively purer than in the natural environment. Compared to the natural level this level should be at least 2-5 fold greater (e.g., in terms of mg/mL). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of 10 magnitude is expressly contemplated. The substance is preferably free of substances present in its natural environment at a functionally significant level, for example 90%, 95%, or 99% pure.

In preferred embodiments, the polypeptide is a fragment of 15 the protein encoded by the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18. Preferably, the Yia operon polypeptide contains at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous 20 amino acids set forth in the full-length amino acid sequence of SEQ ID NO:10; at least 30, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:11; at least 5, 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth 25 in the full-length amino acid sequence of SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14; at least 17, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length amino acid sequence of SEQ ID NO:15, SEQ ID NO:17, or SEQ ID NO:18; at least 11, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids set forth in the full-length 30 amino acid sequence of SEQ ID NO:16, or a functional derivative thereof.

The polypeptide can be isolated from a natural source by methods well-known in the art. The natural source may be 35 protozoal, eukaryotic, or prokaryotic, and the polypeptide may be

synthesized using an automated polypeptide synthesizer. Preferably, the polypeptide is isolated, enriched, or purified from bacteria, most preferably *Klebsiella oxytoca*.

In some embodiments the invention includes one or more recombinant Yia operon-related polypeptides. By "recombinant Yia operon-related polypeptide" is meant a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (e.g., present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

In a sixth aspect, the invention features an antibody (e.g., a monoclonal or polyclonal antibody) having specific binding affinity to a Yia operon-related polypeptide or a Yia operon-related polypeptide fragment. In preferred embodiments, the yia operon-related polypeptide is selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

By "specific binding affinity" is meant that the antibody binds to the target Yia operon-related polypeptide with greater affinity than it binds to other polypeptides under specified conditions. Antibodies or antibody fragments are polypeptides which contain regions that can bind other polypeptides. The term "specific binding affinity" describes an antibody that binds to a Yia operon polypeptide with greater affinity than it binds to other polypeptides under specified conditions.

The term "polyclonal" refers to antibodies that are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen or an antigenic functional derivative thereof. For the production of polyclonal antibodies, various host animals may be immunized by injection with the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species.

"Monoclonal antibodies" are substantially homogenous populations of antibodies to a particular antigen. They may be

obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. Monoclonal antibodies may be obtained by methods known to those skilled in the art (Kohler et al., Nature 256:495-497, 1975, and 5 U.S. Patent No. 4,376,110, both of which are hereby incorporated by reference herein in their entirety including any figures, tables, or drawings).

The term "antibody fragment" refers to a portion of an antibody, often the hypervariable region and portions of the 10 surrounding heavy and light chains, that displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an antibody that physically binds to the polypeptide target.

Antibodies or antibody fragments having specific binding 15 affinity to a Yia operon-related polypeptide of the invention may be used in methods for detecting the presence and/or amount of Yia operon polypeptide in a sample by probing the sample with the antibody under conditions suitable for Yia operon-related-antibody immunocomplex formation and detecting the presence 20 and/or amount of the antibody conjugated to the Yia operon-related polypeptide. Diagnostic kits for performing such methods may be constructed to include antibodies or antibody fragments specific for the Yia operon-related polypeptide as well as a conjugate of a binding partner of the antibodies or the 25 antibodies themselves.

An antibody or antibody fragment with specific binding affinity to a Yia operon-related polypeptide of the invention can be isolated, enriched, or purified from a prokaryotic or eukaryotic organism. Routine methods known to those skilled in 30 the art enable production of antibodies or antibody fragments, in both prokaryotic and eukaryotic organisms. Purification, enrichment, and isolation of antibodies, which are polypeptide molecules, are described above.

Antibodies having specific binding affinity to a Yia operon- 35 related polypeptide of the invention may be used in methods for

detecting the presence and/or amount of Yia operon-related polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated 5 to the Yia operon-related polypeptide. Diagnostic kits for performing such methods may be constructed to include a first container containing the antibody and a second container having a conjugate of a binding partner of the antibody and a label, such as, for example, a radioisotope. The diagnostic kit may 10 also include notification of an FDA approved use and instructions therefor.

In a seventh aspect, the invention features a hybridoma that produces an antibody having specific binding affinity to a Yia operon-related polypeptide or a Yia operon-related polypeptide fragment. By "hybridoma" is meant an immortalized cell line that is capable of secreting an antibody, for example an antibody to a Yia operon-related polypeptide of the invention. In preferred embodiments, the antibody to the Yia operon-related polypeptide comprises a sequence of amino acids that is able to specifically 20 bind a Yia operon-related polypeptide of the invention.

In an eighth aspect, the invention features a Yia operon-related polypeptide binding agent able to bind to a Yia operon-related polypeptide. The binding agent is preferably a purified antibody that recognizes an epitope present on a Yia operon-related polypeptide of the invention. Other binding agents 25 include molecules that bind to Yia operon-related polypeptides and analogous molecules which bind to a Yia operon-related polypeptide. Such binding agents may be identified by using assays that measure Yia operon-related binding partner activity, 30 such as those that measure growth or ascorbate metabolism.

The invention also features a method for screening for other organisms containing a Yia operon-related polypeptide of the invention or an equivalent sequence. The method involves identifying the novel polypeptide in other organisms using 35 techniques that are routine and standard in the art, such as

those described herein for identifying the Yia operon-related polypeptide of the invention or others standard in the art (e.g., cloning, Southern or Northern blot analysis, in situ hybridization, PCR amplification, etc.).

5 A ninth aspect of the invention features a method for identifying a substance that converts a source compound to a target compound, comprising: contacting a cell with nucleic acid, where the nucleic acid expresses a product that converts a source compound into a target compound, and where the cell
10 expresses one or more proteins which in the presence of the target compound provide a detectable signal; contacting the cell with a test substance; and monitoring the detectable signal, where the detectable signal indicates the presence of the substance.

15 In preferred embodiments of the method for identifying a substance that converts a source compound to a target compound, the substance is selected from the group consisting of antibodies, small organic molecules, peptidomimetics, and natural products. In other preferred embodiments, the detectable signal
20 is selected from a group consisting of growth, fluorescence, luminescence, and color. Preferably, the detectable signal is growth, and the target compound is metabolizable to an element selected from the group consisting of carbon, nitrogen, sulfur, and phosphorous, most preferably carbon. Alternatively, the
25 target compound is metabolizable to an essential nutrient. In still other preferred embodiments of the invention, the source compound is selected from the group consisting of 2-KLG, 2,5-DKG, L-IIA, L-GuA, and glucose.

30 In other highly preferred embodiments of the method for identifying a substance that converts a source compound to a target compound, the one or more proteins are one or more Yia operon-related polypeptides. Preferably, the Yia operon further comprises a vector or promoter effective to initiate transcription in a host cell, and most preferably the vector or

promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene.

A tenth aspect of the invention features a method for detecting the presence, absence, or amount of a compound in a sample comprising: contacting the sample with a cell, where the cell expresses one or more genes encoding one or more proteins that in the presence of the compound provide a detectable signal that indicates the presence, absence, or amount of said compound. A schematic of an example of a preferred embodiment of the method is shown in Fig. 13. In preferred embodiments, the compound is ascorbate and the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color. In other preferred embodiments, the one or more genes comprises *yiaJ*, and preferably further comprises a promoter transcriptionally linked to a reporter gene. Preferably, *YiaJ* is naturally expressed in the cell, or the cell has been genetically manipulated to express *YiaJ*. Preferably the reporter gene has a promoter transcriptionally linked and the expression of the reporter gene is regulated by the binding of *YiaJ* to the promoter. The binding of *YiaJ* to the promoter is preferably regulated by the presence or absence of ascorbate. Preferably the cell is a bacteria, and most preferably *Klebsiella oxytoca*.

An eleventh aspect of the invention features an isolated, purified, or enriched nucleic acid molecule encoding *YiaJ* and a reporter gene. Preferably, the nucleic acid molecule further comprises a promoter transcriptionally linked to a reporter gene. Preferably the reporter gene is regulated by the binding of *YiaJ* to the promoter. The binding of *YiaJ* to the promoter is preferably regulated by the presence or absence of ascorbate. In preferred embodiments, the nucleic acid molecule further comprises a vector or promoter effective to initiate transcription in a host cell.

A twelfth aspect of the invention features a recombinant cell comprising the nucleic acid molecule described in the eleventh aspect of the invention, above.

Preferred embodiments of this aspect of the invention feature a recombinant cell for detecting the presence, absence, or amount of a compound in a sample, where the cell expresses one or more genes encoding one or more proteins that in the presence 5 of the compound provide a detectable signal, where the signal indicates the presence, absence, or amount of the compound. In preferred embodiments, the detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

10 In other preferred embodiments of the recombinant cell for detecting the presence, absence, or amount of a compound in a sample, the one or more genes comprises *yiaJ*, and further comprises a promoter transcriptionally linked to a reporter gene. Preferably, the expression of the reporter gene is regulated by 15 the binding of *YiaJ* to the promoter. Preferably, *yiaJ* is naturally expressed in the recombinant cell, or the cell has been genetically manipulated to express *yiaJ*. The recombinant cell is preferably a bacteria, and more preferably *Klebsiella oxytoca*.

A thirteenth aspect of the invention features a method of 20 selection for one or more nucleic acid sequences encoding a metabolic pathway from a source compound to a target compound comprising: (1) identifying an organism that metabolizes a target compound to provide an essential element; (2) identifying one or more genes responsible for the metabolism of the target compound 25 to the essential element; (3) expressing the one or more genes under the control of an inducible promoter, whereby the target compound is metabolized only in the presence of an inducer and not in the absence of the inducer; (4) expressing nucleic acid sequences potentially encoding the metabolic pathway in the 30 recipient organism; and (5) selecting the recipient organism for growth in the presence of the source compound in the absence of the target compound and in the presence of the inducer, where growth on the source compound in the absence of the target compound and in the presence of the inducer indicates the 35 presence of the nucleic acid sequence.

In preferred embodiments of the method of selection, the essential element is selected from the group consisting of carbon, phosphorous, nitrogen, and sulfur, and most preferably is carbon.

5 In other preferred embodiments, the method of selection further comprises the transfer of the one or more genes to a highly genetically manipulatable recipient organism, such that the recipient organism metabolizes the target compound to provide an essential element.

10 By a "highly genetically manipulatable recipient organism" is meant an organism, preferably single-celled, more preferably bacteria, and most preferably *Klebsiella oxytoca*, that can be manipulated by the standard genetic techniques, including but not limited to, transfection, selection in selective media, growth in culture.

15 The summary of the invention described above is not limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

20 Description Of The Figures

Figure 1 shows a physical map of the *yiaK-S* operon, which includes the open reading frames *yiaK*, *yiaL*, *orf1*, *yiaX2*, *lyxK*, *yiaQ*, *yiaR*, and *yia*, and its putative regulator, *yiaJ*, compared with the *E. coli* *yiaK-S* operon, which includes the open reading frames *yiaK*, *yiaL*, *yiaM*, *yiaN*, *yiaO*, *lyxK*, *yiaQ*, *yiaR*, and *yiaS*, and its putative regulator *yiaJ*.

25 Figures 2A, 2B, 2C, 2D, 2E, and 2F show the nucleic acid sequence (SEQ ID NO:19) and translated amino acid sequences of the open reading frames of the *yia* operon and its putative regulator, *yiaJ*.

30 Figure 3 shows a multiple sequence alignment of YiaJ-Ko (SEQ ID NO:10), YiaJ-Ec (SEQ ID NO:20), and YiaJ-Hi (SEQ ID NO:21). Identical sequences among the three proteins are indicated by shading.

Figure 4 shows a multiple sequence alignment of YiaK-Ko (SEQ ID NO:11), YiaK-Ec (SEQ ID NO:22), and YiaK-Hi (SEQ ID NO:23). Identical sequences among the three proteins are indicated by shading.

5 Figure 5 shows a multiple sequence alignment of YiaJ-Ko (SEQ ID NO:12), YiaL-Ec (SEQ ID NO:24), and YhcH-Hi (SEQ ID NO:25). Identical sequences among the three proteins are indicated by shading.

10 Figure 6 shows a multiple sequence alignment of LyxK-Ko (SEQ ID NO:15), LyxK-Ec (SEQ ID NO:26), and LyxK-Hi (SEQ ID NO:27). Identical sequences among the three proteins are indicated by shading.

15 Figure 7 shows a multiple sequence alignment of YiaQ-Ko (SEQ ID NO:16), YiaQ-Ec (SEQ ID NO: 28), and YiaQ-Hi (SEQ ID NO:29). Identical sequences among the three proteins are indicated by shading.

20 Figure 8 shows a multiple sequence alignment of YiaR-Ko (SEQ ID NO:17), YiaR-Ec (SEQ ID NO:30), and YiaR-Hi (SEQ ID NO:31). Identical sequences among the three proteins are indicated by shading.

Figure 9 shows a multiple sequence alignment of YiaS-Ko (SEQ ID NO:18), YiaS-Ec (SEQ ID NO:32), and YiaS-Hi (SEQ ID NO:33). Identical sequences among the three proteins are indicated by shading.

25 Figure 10 shows a schematic of the construction of the Tester Strain. The plasmid pMG125 is shown which comprises: (i) a chloramphenicol resistance marker (*cat*); (ii) the thermosensitive origin of replication from plasmid pH01 (*pH01 rep (t')*); (iii) a 0.8 kb fragment containing the 5' region of the *yiaJ* gene 30 and its promoter sequences; (iv) the spectinomycin resistance marker (*spc*); (v) the *lacI^q-lacO-trc* promoter fragment; and (vi) a 1 kb fragment containing the 5' end of *yiaK*, including its ribosome binding site for translation initiation while excluding the promoter sequences of the *yiaK-S* operon. The recombinant 35 plasmid pMG125 was introduced into *K. oxytoca* wild type strain

VJSK009 by transformation at 30 °C, the permissive temperature for pMAK705 replication. Chromosomal integration of the pMG125 insert into VJSK009 was achieved by double crossover at the *yiaJ-K* locus such that the endogenous promoter of the *yiaK-S* operon 5 was replaced with the inducible *lacI^q-trc* promoter system in the resulting recombinant cell, MGK003.

Figure 11 shows a schematic representation of a general example of a metabolic selection process. Briefly, genetic material, isolated from microbes, is incorporated into a Tester 10 Strain and the gene(s) of interest selected for by growth on "S". The gene(s) of interest will catalyze the conversion of "S" to "T" in the Tester Strain, thereby allowing growth on "S".

Figure 12 shows a schematic representation of a more specific example of metabolic selection process, in which "S" is 15 2-KLG and "T" is AsA. In this case, the gene(s) of interest are those that catalyze the conversion of 2-KLG to AsA.

Figure 13, part A shows a theoretical model for AsA-dependent activation of the *yiaK-S* operon. Based on transcriptional analyses, the YiaJ regulatory protein is thought 20 to activate transcription of the *yiaK-S* AsA catabolic operon in response to AsA present in the medium. However, the inventors do not wish to be held to this interpretation of the data.

Figure 13, part B shows a schematic representation of a whole-cell reporter system for AsA sensing. The *yiaK-S* promoter 25 region (P_{yia}) is fused to the Green-Fluorescent-Protein (GFP) gene (or to *lux* or other reporter genes), and the fusion is integrated into the chromosome of an indicator strain, which also contains the YiaJ regulator. In the presence of AsA, YiaJ is stimulated and activates transcription of the *yia-GFP* fusion, thereby 30 conferring an easily detectable GFP-positive or fluorescent phenotype.

Detailed Description Of The Invention

The instant invention is based in part on the use of a metabolic selection strategy that uses a recombinant DNA

selection procedure to identify enzymatic pathways for the conversion of a source compound to a target compound. This technique allows at least a million-fold increase in the discovery rate over classical biochemical screening approaches, 5 and allows testing of the 99% of the environmental microbes that are currently not able to be cultured in the laboratory.

The general process involves the creation/ identification of an easily genetically-manipulatable organism containing an inducible signal, such that the signal is activated when a target 10 compound is metabolized, followed by the screening of nucleic acid in this organism to identify genes which metabolize a source compound to the target compound (Figs. 11 and 12)

In a specific embodiment, the process involves three steps (1) the identification of an organism capable of metabolizing the target compound to carbon and energy, and the transfer of this metabolic pathway to a highly genetically manipulatable organism, e.g. *Escherichia coli* or *Bacillus subtilis*, with the result that the recipient now uses the target compound for growth; (2) placing the expression of the pathway under the control of an inducible promoter, whereby the target compound is metabolized in the presence of an inducer and not in its absence; and (3) cloning genes, which are to be tested for their ability to metabolize the source compound, into the recipient, and selecting for growth on the source compound in the presence of the inducer 25 but in the absence of the target compound.

Once positive organisms are identified in the above selection scheme by growth in the presence of inducer, the organisms are further screened for their ability to grow in the absence of the inducer. No growth in the absence of the inducer 30 indicates that the metabolism of the source compound proceeds via the target compound. Thus, the nucleic acid probably encodes an enzymatic pathway for the conversion of the source compound to the target compound.

Growth in the absence of the inducer indicates that 35 metabolism of the source compound to the essential element or

factor does not require prior conversion to the target compound, rather it may proceed directly, or through an intermediate, to the essential element or factor. When conversion directly to the target compound is the desired result, further work is necessary
5 to obtain the desired genes. methods of obtaining the desired genes include: re-selection of DNA from other sources; random mutation of the DNA followed by re-selection; knocking out (deleting or blocking the expression of genes by methods well-known in the art) the genes that allow the direct conversion to
10 the essential element or factor or from an intermediate to the essential element or factor followed by re-selection; etc. In one preferred embodiment, expression of the genes that allow the direct, or partially direct, conversion to the essential factor are knocked out or their expression blocked, thereby "forcing"
15 the conversion to the essential element through the target compound. This will be effective if a pathway through the target compound existed, but was thermodynamically unfavorable, for example.

Alternatively, if the intermediate is freely
20 interconvertable with the desired target compound as well as to the essential element, growth in the absence of the inducer may be an acceptable outcome, or even desirable. By "freely interconvertable" is meant that an enzymatic pathway is present to allow the intermediate to be converted to the target. The
25 interconvertability of the compounds would also be determined using the methods described above for obtaining a pathway directly to the target compound.

Under some circumstances, selection of a pathway directly, or through an intermediate, to the essential element or factor
30 rather than to the target compound, is a preferred result. For example, under circumstances where the desired target compound is not one that can be used for direct selection (e.g. does not cross membranes or is rapidly broken down) a "surrogate target" might have to be used. A surrogate target refers to one that is
35 used for selection, but is not the most highly desired target. In

this embodiment, the target would preferably be on the pathway of conversion of the surrogate target to the essential element.

I.. Functional Derivatives

Provided herein are functional derivatives of a polypeptide or nucleic acid of the invention. By "functional derivative" is meant a "chemical derivative," "fragment," or "variant," of the polypeptide or nucleic acid of the invention, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with an antibody specific for the protein, enzymatic activity or binding activity mediated through noncatalytic domains, which permits its utility in accordance with the present invention. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide which retains the functionality of the original. In both cases, all permutations are intended to be covered by this disclosure.

Also included with "functional derivatives" of the polypeptides, in particular, of the invention are "chemical derivatives". A "chemical derivative" contains additional chemical moieties not normally a part of the protein. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, for example, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate,

N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing primary amine containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4-pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyltyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimide ($R'-N-C-N-R'$) such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residue are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutamyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope 5 of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking the component peptides of the protein to each other or to other proteins in a complex to a water-insoluble support matrix or to other macromolecular carriers. 10 Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and 15 bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl] dithiol-propioimide yield photo-activatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen 20 bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and 25 lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal 30 amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half-life, and the like. The moieties may alternatively eliminate or attenuate any undesirable 35 side effect of the protein complex and the like. Moieties

capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the proteins, of the complexes having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence.

Fragments of a protein are useful for screening for compounds that act to modulate enzyme activity, as described herein. It is understood that such fragments may retain one or more characterizing portions of the native complex. Examples of such retained characteristics include: catalytic activity; substrate specificity; interaction with other molecules in the intact cell; regulatory functions; or binding with an antibody specific for the native complex, or an epitope thereof.

Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring complex component by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. It is understood that such variants having added, substituted and/or additional amino acids retain one or more characterizing portions of the native protein, as described above.

A functional derivative of a protein with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the

art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman et al., 1983, DNA 2:183) wherein nucleotides in the DNA coding the sequence are modified, 5 and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, proteins with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art. 10 The functional derivatives of the proteins typically exhibit the same qualitative biological activity as the native proteins.

II. Nucleic Acid Probes, Methods, and Kits for Detection of Yia operon-related polypeptides

A nucleic acid probe of the present invention may be used to 15 probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain other nucleic acid molecules of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (cf. "Molecular Cloning: A Laboratory Manual", second 20 edition, Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989).

In the alternative, chemical synthesis can be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the 25 amino acid sequence of the polypeptide of interest. The synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", Academic Press, 30 Michael, et al., eds., 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer

alignment and sequence analysis known in the art ("Molecular Cloning: A Laboratory Manual", 1989, supra). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, 5 fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA, as well as DNA probes, such probes being generated using 10 techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for 15 coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used 20 in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the 25 method utilized.

One method of detecting the presence of nucleic acids of the invention in a sample comprises (a) contacting said sample with the above-described nucleic acid probe under conditions such that hybridization occurs, and (b) detecting the presence of said 30 probe bound to said nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples extracted from environmental samples.

A kit for detecting the presence of nucleic acids of the invention in a sample comprises at least one container means having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or 5 more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymatic labeled probes (horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, 10 avidin, or streptavidin). Preferably, the kit further comprises instructions for use.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of 15 plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such 20 containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the 25 hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

30 III. DNA Constructs Comprising Yia Operon-Related Nucleic Acid Molecules and Cells Containing These Constructs.

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid

molecules. In addition, the present invention relates to a recombinant DNA molecule comprising a vector and an above-described nucleic acid molecule. The present invention also relates to a nucleic acid molecule comprising a transcriptional 5 region functional in a cell, a sequence complementary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

10 The present invention also relates to a cell or organism that contains an above-described nucleic acid molecule and thereby is capable of expressing a polypeptide. The polypeptide may be purified from cells which have been altered to express the polypeptide. A cell is said to be "altered to express a desired 15 polypeptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either 20 eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to 25 nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may 30 vary from organism to organism, but shall in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 35 5'-non-coding sequences involved with initiation of transcription

and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a Yia operon polypeptide of the invention may be obtained by the above-described methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a polypeptide of the invention, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.

Two DNA sequences (such as a promoter region sequence and a sequence encoding a polypeptide of the invention) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a gene sequence encoding a polypeptide of the invention, or (3) interfere with the ability of the gene sequence of a polypeptide of the invention to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a gene encoding a polypeptide of the invention, transcriptional and translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a gene encoding a polypeptide of the invention (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for polypeptides of the invention. Prokaryotes most frequently are represented by

various strains of *E. coli*. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC18, pUC19 and the like; suitable phage or bacteriophage vectors may include γ gt10, γ gt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, *Klebsiella*, and the like. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express a polypeptide of the invention (or a functional derivative thereof) in a prokaryotic cell, it is necessary to operably link the sequence encoding the polypeptide of the invention to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (i.e., inducible or derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage λ , the *bla* promoter of the β -lactamase gene sequence of pBR322, and the *cat* promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage λ (P_L and P_R), the *trp*, *recA*, *lacZ*, *lacI*, and *gal* promoters of *E. coli*, the α -amylase (Ulmanen et al., J. Bacteriol. 162:176-182, 1985) and the ζ -28-specific promoters of *B. subtilis* (Gilman et al., Gene Sequence 32:11-20, 1984), the promoters of the bacteriophages of *Bacillus* (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY, 1982), and *Streptomyces* promoters (Ward et al., Mol. Gen. Genet. 203:468-478, 1986). Prokaryotic promoters are reviewed by Glick

(Ind. Microbiol. 1:277-282, 1987), Cenatiempo (Biochimie 68:505-516, 1986), and Gottesman (Ann. Rev. Genet. 18:415-442, 1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome-binding site upstream of the gene sequence-encoding sequence. Such ribosome-binding sites are disclosed, for example, by Gold et al. (Ann. Rev. Microbiol. 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene. As used herein, "cell", "cell line", and "cell culture" may be used interchangeably and all such designations include progeny. Thus, the terms "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. However, as long as mutant progeny have the same functionality as that of the originally transformed cell, they are considered to be the same cell or cell-line.

Host cells which may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the polypeptide of interest. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

A nucleic acid molecule encoding a polypeptide of the invention and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a nonreplicating DNA or RNA molecule, which may either be a linear molecule or a closed covalent circular molecule. Alternatively, permanent expression may occur through the integration of the

introduced DNA sequence into the host chromosome or as a circular plasmid.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells 5 which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy 10 metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal 15 synthesis of mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama (Mol. Cell. Biol. 3:280-289, 1983).

The introduced nucleic acid molecule can be incorporated 20 into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and 25 selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those 30 capable of replication in *E. coli* (such as, for example, pBR322, ColE1, pSC101, pACYC 184, π VX; "Molecular Cloning: A Laboratory Manual", 1989, supra). *Bacillus* plasmids include pC194, pC221, pT127, and the like (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, NY, pp. 307-329, 1982). Suitable 35 *Streptomyces* plasmids include pIJ101 (Kendall et al., J.

Bacteriol. 169:4177-4183, 1987), and streptomyces bacteriophages such as φC31 (Chater et al., In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kaido, Budapest, Hungary, pp. 45-54, 1986). *Pseudomonas* plasmids are reviewed by John et al. 5 (Rev. Infect. Dis. 8:693-704, 1986), and Izaki (Jpn. J. Bacteriol. 33:729-742, 1978).

Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by 10 any of a variety of suitable means, i.e., transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which 15 selects for the growth of vector-containing cells. Expression of the cloned gene(s) results in the production of a polypeptide of the invention, or fragments thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of 20 bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

25 V. Antibodies, Hybridomas, Methods of Use and Kits for Detection of Yia Operon-Related polypeptides

The present invention relates to an antibody having binding affinity to a polypeptide of the invention. The polypeptide may have the amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:10, 30 SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, or a functional derivative thereof, or at least 6 contiguous amino acids thereof (preferably, at least 15, 20, 25, 30, 35, or 40 contiguous amino acids thereof).

The present invention also relates to an antibody having specific binding affinity to a polypeptide of the invention. Such an antibody may be isolated by comparing its binding affinity to a polypeptide of the invention with its binding 5 affinity to other polypeptides. Those which bind selectively to a polypeptide of the invention would be chosen for use in methods requiring a distinction between a polypeptide of the invention and other polypeptides. Such methods could include, but should not be limited to, the identification of other cells expressing 10 the polypeptides of the invention.

The polypeptides of the present invention can be used in a variety of procedures and methods, such as for the generation of antibodies, for use in identifying pharmaceutical compositions, and for selection of other enzymmatic pathways.

15 The polypeptides of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide could be generated as described herein and used as an immunogen. The antibodies of the present invention include monoclonal and 20 polyclonal antibodies, as well fragments of these antibodies.

The present invention also relates to a hybridoma which produces the above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

25 In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1984; St. Groth et al., J. Immunol. Methods 35:1-21, 30 1980). Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the 35 amount of polypeptide used for immunization will vary based on

the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide may be modified or administered in an adjuvant in order to increase the peptide antigenicity. Methods 5 of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or β -galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized 10 animals are removed, fused with myeloma cells, such as SP2/0-Agl4 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These 15 include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz et al., *Exp. Cell Res.* 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass are determined using procedures known in the art (Campbell, "Monoclonal Antibody 20 Technology: Laboratory Techniques in Biochemistry and Molecular Biology", *supra*, 1984).

For polyclonal antibodies, antibody-containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of 25 the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horse radish peroxidase, alkaline phosphatase, and the like) 30 fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, see Stemberger et al., *J. Histochem. Cytochem.* 18:315, 1970; Bayer et al., *Meth. Enzym.* 62:308-, 1979; Engval et al., *Immunol.* 35 109:129-, 1972; Goding, *J. Immunol. Meth.* 13:215-, 1976. The

labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues which express a specific peptide.

The above-described antibodies may also be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby et al., Meth. Enzym. 34, Academic Press, N.Y., 1974). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as immuno-chromatography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed herein with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides (Hurby et al., "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307, 1992; Kaspczak et al., Biochemistry 28:9230-9238, 1989).

Anti-peptide peptides can be generated by replacing the basic amino acid residues found in the peptide sequences of the Yia operon polypeptides of the invention with acidic residues, while maintaining hydrophobic and uncharged polar groups. For example, lysine, arginine, and/or histidine residues are replaced with aspartic acid or glutamic acid and glutamic acid residues are replaced by lysine, arginine or histidine.

The present invention also encompasses a method of detecting a Yia operon-related polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail,

the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying whether the antibody binds to the test sample. Detection of a polypeptide of the invention in a sample may indicate the presence of the 5 pathway of the invention in other cells.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will 10 recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion-based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can 15 be found in Chard ("An Introduction to Radioimmunoassay and Related Techniques" Elsevier Science Publishers, Amsterdam, The Netherlands, 1986), Bullock et al. ("Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1, 1982; Vol. 2, 1983; Vol. 3, 1985), Tijssen ("Practice and Theory of 20 Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, 25 or environmental samples. The test samples used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can 30 readily be adapted in order to obtain a sample which is testable with the system utilized.

A kit contains all the necessary reagents to carry out the previously described methods of detection. The kit may comprise: 35 (i) a first container means containing an above-described antibody, and (ii) second container means containing a conjugate

- comprising a binding partner of the antibody and a label. Preferably, the kit also contains instructions for use. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following:
- 5 wash reagents and reagents capable of detecting the presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or 10 antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit may be as described above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established 15 kit formats which are well known in the art.

Other methods associated with the invention are described in the examples disclosed herein.

Examples

The examples below are not limiting and are merely 20 representative of various aspects and features of the present invention. The examples below demonstrate the construction and use of metabolic selection systems, and the isolation of desired enzymatic pathways.

EXAMPLE 1: Construction of a Tester Strain for the Selection of Pathways from 2-KLG to AsA

This example is exemplary of how to construct tester strains, and therefore can be applied to the identification and construction of tester strains for the selection of other metabolic pathways. The basic idea is to take environmental 30 samples and test them for growth on a target compound (in the example, ascorbate). Then, positive colonies are screened for the inability to grow on the source compound (in the example, 2-KLG). The tester strain is the one that grows on the target, but

not the source compound. Once the genes encoding the metabolic pathway for the target compound to the essential factor (an element such as carbon, nitrogen, sulphur or phosphorous, or a nutrient, for example) are identified, they are then placed under 5 the control of an inducible promoter, and the tester strain is ready to be utilized to select for the metabolic pathway from the source to the target compound.

If it proves difficult to obtain a tester strain that grows on the target, but not the source, but strains exist that do not 10 grow on the source, then the pathway that permits growth on the target can be isolated and transferred to another strain that doesn't grow on the source in order to obtain the desired tester strain.

Isolation of a Strain that Grows on AsA, but not 2-KLG

15 Samples from diverse natural environments were collected to use for the isolation of microbes that can utilize ascorbic acid (AsA) as the sole carbon source. No bacterial species has previously been reported to grow on AsA minimal medium.

Environmental samples were collected from freshwater lakes, 20 lemon and orange orchards, residential backyard soils, human and animal solid wastes.

Over 100 microbial isolates, capable of forming visible colonies within 20 hours of incubation at 30 °C on M9 minimal medium containing 0.5% AsA, were selected from these samples. 25 These 100 isolates were then screened for their ability to grow on 2-Keto-L-Gulonate (2-KLG) minimal medium.

One of the isolates that could utilize AsA as its sole source of carbon and energy, but could not grow on 2-KLG, was identified as *Klebsiella oxytoca* (Table 1). Thus, *Klebsiella oxytoca* was retained as a candidate for genetic engineering of a host strain that can use AsA under controlled conditions for the selection of cloned microbial pathways from 2-KLG to AsA.

Other bacterial strains capable of metabolizing ascorbic acid to carbon and energy were also identified, as were some that also metabolized 2KLG to carbon and energy (Table 1).

Table 1

5

COMPOUND UTILIZATION OF ENVIRONMENTAL ISOLATES

	AsA	2-KLG
<u>GRAM POSITIVES</u>	<u>72 HR</u>	<u>24 HR</u>
<i>Bacillus megaterium</i>	+	+
<i>Streptomyces species</i>	++	++
<i>Yellow Bug</i>	++	+++
<u>GRAM NEGATIVES</u>	<u>24 HR</u>	<u>72 HR</u>
<i>Klebsiella pneumoniae</i>	+++	-
<i>Klebsiella species</i>	+++	-
<i>Klebsiella oxytoca</i>	+++	-
<i>Unknown Malodorous</i>	++	-
<i>Short Rod</i>		

Identification of Genes Responsible for AsA Catabolism

In order to identify the gene(s) responsible for AsA catabolism in *K. oxytoca*, mutagenesis by transposition insertion 10 was performed in *K. oxytoca* strain VJSK009 (Cali, B. M., et al., 1989. J. Bacteriol. 171:2666-2672) using the pfd-Tn5 delivery vector as described by Metzger, M., et al., 1992. Nucl. Acids Res. 20:2265-2270. Among 5,000 clones screened, several mutants 15 that were no longer capable of growing on AsA were identified, most of which were also affected in their ability to grow on conventional carbon sources such as glucose, maltose, pyruvate or succinate. Two of the mutants, however, were specifically affected in AsA utilization and were further characterized by cloning and sequencing the regions adjacent to the transposon 20 insertion.

Characterization of the Genes/Proteins of the Operon

In both mutants, the Tn5 insertion was found to disrupt the same operon of 8 genes. This operon was found to be homologous to the *yiaK-S* operon of *E. coli* (Blattner, F. R., et al., 1997.

- 5 Science 277:1453-1462) which is thought to be involved with carbohydrate utilization (Badia, J., et al., 1998. J. Biol. Chem. 273:8376-8381).

Similarly to *E. coli*, the *K. oxytoca* *yiaK-S* operon is preceded by a transcriptional regulator, *yiaJ*. A physical map of 10 the *yiaK-S* operon and its putative regulator is shown in Figure 1. The nucleic acid sequence and translated amino acid sequence of the open reading frames of the operon and its putative regulator are shown in Figure 2 A-F.

The functions of the *yia* operon gene products in *K. oxytoca* 15 and *E. coli* are unknown, except for the *E. coli* *lyxK*-encoded enzyme which was shown to phosphorylate L-xylulose and play a key role in the utilization of L-lyxose by *E. coli* (Sanchez, J. C., et al., 1994. J. Biol. Chem. 169:29665-29669). However, the *yiaK-S* operon is thought to be silent in wild-type *E. coli*, L- 20 xylulose activity could not be detected in wild type cells, and *E. coli* K12 is unable to metabolize L-lyxose (Sanchez, J. C., et al., 1994. *supra*). A similar operon is also present in *Haemophilus influenzae*, but no function has been determined for any of the open reading frames (Fleischmann, R.D., et al., 1995. 25 Science 269:496-512).

Alignments of the *yia* open reading frames common among the three species are shown (Figs. 3-9). Based on sequence 30 similarities, *yiaQ* has been classified as a putative hexulose-6-phosphate synthase, *yiaR* as a putative hexulose-6-phosphate isomerase, and *yiaS* as a putative sugar isomerase (data not shown).

Place Operon under the control of an Inducible Promoter

To engineer *K. oxytoca* as a host strain for the selection of biocatalysts which produce AsA, the promoter of the *yiaK-S* operon

was replaced with a DNA fragment that contained the *trp-lac* hybrid promoter of transcription, the *lacO* operator, and the *lacI^q* repressor gene (Brosius, J. 1992. *Meth. Enzymol.* **216**:469-483). This allows the *yiaK-S* operon, and therefore AsA catabolism, to be turned ON and OFF in a tightly controlled manner in the presence or absence of IPTG, a non-metabolizable inducer of the *lac* promoter. Practically, a 5-way ligation was set up among: (i) the pMAK705 integration vector which carries a chloramphenicol resistance marker and the thermosensitive origin of replication from plasmid pH01 (Hamilton, C. M., et al., 1989. *J. Bacteriol.* **171**:4617-4622); (ii) a 0.8 kb fragment containing the 5' region of the *yiaJ* gene and its promoter sequences; (iii) the spectinomycin resistance marker retrieved from *Staphylococcus aureus* Tn554 (Murphy, E. 1985. *Mol. Gen. Genet.* **200**:33-39) to follow integration events; (iv) the *lacI^q-lacO-trc* promoter fragment retrieved from pSE380 (InVitrogen, Carlsbad, CA); and (v) a 1 kb fragment containing the 5' end of *yiaK*, including its ribosome binding site for translation initiation while excluding the promoter sequences of the *yiaK-S* operon (Figure 10).

The recombinant plasmid, pMG125, was introduced into *K. oxytoca* wild type strain VJSK009 by transformation at 30 °C, the permissive temperature for pMAK705 replication. Chromosomal integration of the pMG125 insert by double crossover at the *yiaJ-K* locus was achieved by successive temperature switches as described by (Hamilton, C. M., et al., 1989. *supra*). PCR analyses were performed on 12 candidates to verify that the endogenous promoter of the *yiaK-S* operon had been replaced with the inducible *lacI^q-trc* promoter system (Figure 10).

The resulting strain, MGK003, proved able to grow on M9 minimal medium supplemented with AsA 0.25% and IPTG 10 to 100 µM, while no growth was observed on the same medium lacking IPTG.

EXAMPLE 2: Preparation of Environmental DNA Libraries

An example of a currently preferred method for the isolation of DNA from environmental samples is provided below. In the

example, purification from soil and water samples are described, however samples can be from any environmental source and the methods adapted according to practices well-known in the art.

Direct Isolation of Total DNA from Soil and Water Samples

- 5 Total microbial DNA was isolated from various soil and water samples according to the following procedure which is derived and modified from Steffan, R.J., et al., 1988. Appl. Environ. Microbiol. **54**:2908-2915; Whatling, C. A., and C. M. Thomas. 1993. Anal. Biochem. **210**:98-101; and Zhou, J., et al., 1996. Appl. Environ. Microbiol. **62**:316-322.
- 10 1. Begin with 100 g wet soil or 50 g dry soil; 150 mL sodium phosphate buffer 0.1 M, pH 4.5; and 5 g PVPP (acid washed).
- 15 2. Blender - medium speed - 3 times for 1 min (cool down between each cycle). Add 0.5 mL SDS 20%, blend 5 more seconds.
- 20 3. Centrifuge 10 min at 1,000 g at 10 °C.
- 25 4. Keep supernatant. Repeat extraction twice with soil pellet.
- 30 5. Combine the 3 supernatants. Centrifuge 20 min at 10,000 g at 10 °C.
- 35 6. Wash pellet with cold 0.1% sodium-0.1% sodium pyrophosphate. Homogenize with blender for 1 min or shake. Centrifuge 20 min at 10,000 g at 10 °C.
- 40 7. Wash pellet with 33 mM Tris-HCl, 1 mM EDTA, pH 8.0.
- 45 8. Resuspend in 2 mL 10 mM Tris, pH 7.6; 1 N NaCl.
- 50 9. Mix with equal volume 1.2% LMP agarose at 42 °C. Pour into 1 mL syringes. Polymerize for 20 min at 4 °C.
- 55 10. Incubate 3-4 hours at 37 °C in 20 vol. 1 N NaCl; 100 mM EDTA; 10 mM Tris, pH 7.5; 1% sarkosyl; 1 mg/mL lysozyme.
- 60 11. Add 1 mg/mL proteinase K. Incubate overnight at 45 °C.
- 65 12. Wash agarose plugs twice with TE. Store in 100 mM EDTA; 10 mM Tris at 4 °C.

13. Load noodles on LMP agarose gel 0.7%. Cut out chromosomal band. Heat 15 min at 65 °C in TE buffer. Add 2 U GelZyme (InVitrogen) per 200 µL 1% agarose. Incubate for 2 h at 40 °C. EtOH precipitate for no more than 30 min at -20 °C.

5

Preparation of Total DNA from Post-Enrichment Cultures

Aliquots from 18 water or soil samples were used to inoculate 50 mL of M9 minimal medium supplemented with any one of the following carbon sources: 0.5% 2-KLG; 0.25% L-idonate (L-IIA); 10 0.25% L-gulonate (L-GUA) and 0.25% ascorbate. Culture flasks were incubated for 2 to 3 days at 30 °C without agitation.

Total DNA was isolated from these cultures as follows:

1. 20 mL were centrifuged for 5 min at 6,000 rpm.
15 2. Pellets were washed with 5 mL Tris 10 mM, EDTA 1 mM pH 8.0 (TE), were centrifuged again, and were resuspended in 0.9 mL TE.

3. Lysozyme (5 mg/mL) and RNase 100 (µg/mL) were added, and cells were incubated for 10 min at 37 °C.

20 4. Sodium dodecylsulfate (SDS) was added to a final concentration of 1%, and the tubes were gently shaken until lysis was completed.

5. 200 mL of a 5 N NaClO₄ stock solution were added to the lysate.

25 6. The mixture was extracted once with one volume of phenol:chloroform (1:1) and once with one volume of chloroform.

7. Chromosomal DNA was precipitated by adding 2 mL of cold (-20 °C) ethanol and gently coiling the precipitate around a curved Pasteur pipette.

30 8. DNA was dried for 30 min at room temperature and was resuspended in 100 to 500 µL of Tris 10 mM, EDTA 1 mM, NaCl 50 mM pH 8.0 to obtain a DNA concentration of 0.5 to 1 µg/µL.

EXAMPLE 3: Selection for Nucleic Acid which Converts 2-KLG to AsA (Fig. 12)

This example is exemplary of how to select for nucleic acid sequences that encode metabolic pathways, and therefore can be applied to the identification and selection of sequences encoding other metabolic pathways. Basically, a nucleic acid library is made, according to methods well-known in the art, from nucleic acid sequences isolated from environmental samples (as described in Example 2, for example). This library is then transfected into the tester strain and the resulting pool of transfected cells selected for growth on the source compound (2-KLG in the example) in the absence of the target compound (ascorbate in the example) and the presence of the inducer.

Construction of an Enrichment DNA Library in a Cosmid Vector

The SuperCos1 cosmid vector (Stratagene, La Jolla, CA) is a λ -based cloning system suitable for the cloning of large DNA fragments. After treatment according to the manufacturer's instructions, the 8 kb-long vector appears as two arms flanked by cos sites which are recognized by the λ -packaging machinery. Since only DNA molecules from 40 to 48 kb are efficiently packaged in λ -heads, this allows the selective cloning of 32 to 40 kb inserts between the two arms.

Chromosomal DNA extracted from 20 post-enrichment cultures was mixed in equal amounts. Five to ten μ g of the mixture were partially digested with Sau3A restriction enzyme to obtain DNA fragments sized between 5 and 50 kb, were dephosphorylated, and were ligated with SuperCos1 arms using conditions recommended by the supplier. One μ g of the ligation mixture was used in an *in vitro* packaging reaction using the Gigapack III Gold packaging kit from Stratagene to create the cosmid library.

Clearly, this procedure can be used to make other chromosomal DNA libraries, for example from other enriched environmental samples, or from chromosomal DNA extracted directly from environmental samples.

Transfection and Selection of the Cosmid Library

Prior to transfection of *K. oxytoca* strain MGK003 with the packaging mixture, the tester strain was transformed with plasmid pCB382 expressing the *E. coli* *lamB* gene that functions as λ receptor, which appears to be absent or non-functional in most Klebsiella strains (De Vries, G. E., et al., 1984. Proc. Natl. Acad. Sci. USA 81:6080-6084). The resulting MGK003 [λ^s] strain was transfected with the packaged products as follows:

1. Five mL of liquid LB medium supplemented with 0.2% maltose and 10 mM MgSO₄ were inoculated from an overnight preculture of strain MGK003 [pCB382].

2. Cells were grown to an OD₆₀₀ of 0.5, were centrifuged at 500 xg for 10 min, and were resuspended in the same volume of 10 mM MgSO₄.

3. The packaging products were mixed with 2 mL of cells in 15 mL culture tubes, and were incubated for 20 min at 39 °C without shaking.

4. After adding 2.5 mL of 2x YT (1% NaCl; 1% yeast extract; 1.6% tryptone), cells were incubated at 37 °C for 1 h under gentle agitation.

5. A 100 μ L-aliquot was plated on LB-kanamycin medium to determine the number of clones present in the cosmid library.

6. The remainder was centrifuged at 3000 g for 5 min and was resuspended in 1 mL of M9 minimal medium supplemented with 10 μ M IPTG (IPTG concentration can be varied up to 100 μ M), and aliquots (200 μ L) were plated on M9 plates containing 0.5% 2-KLG and 50 μ M IPTG. 7. Plates were incubated at 37 °C for 36 h for selecting candidate pathways that would convert 2-KLG to AsA. (Alternatively, selection can be done at 30 °C.)

30 Among 500,000 clones to which a first selection round was applied, approximately 100 colonies of various sizes appeared on 2-KLG/IPTG plates. These were re-streaked on: (i) LB-kanamycin to verify the presence of the cosmid vector; (ii) 2-KLG/IPTG; and (iii) 2-KLG lacking IPTG to determine if growth of the positive

clones on 2-KLG was dependent upon the expression of AsA catabolism.

Two clones were retained that grew on LB-kanamycin and 2-KLG/IPTG, but not on 2-KLG without IPTG within 20 h at 37 °C. To 5 verify that the observed phenotype was conferred by the cloned DNA, cosmid DNA was extracted from these two clones and introduced, by electroporation, into strain MGK003. In both cases, the back-cross gave a phenotype identical to that of the original clone obtained in the selection process (Data not 10 shown).

Selection of libraries can also be done on other carbon sources to isolate other pathways, for example on L-gulonate (0.25%) plus IPTG to isolate pathways from L-gulonate to AsA, or on L-idonate (0.25%) plus IPTG to isolate pathways from L-idonate 15 to AsA.

EXAMPLE 4: Isolation of Other Pathways

The metabolic selection strategy described above can also be used for the isolation of other pathways of interest, for example from 2-KLG to L-idonate, or 2-KLG to L-gulonate, or 20 alternatively, to identify new reductase enzymes capable of the conversion of 2,5-DKG to 2-KLG. This conversion is one of the slow steps in the production of ascorbate, so identification of an enzymatic method would be economically useful. Basically, the strategy described in the examples above can be used to isolate 25 any pathway to metabolize a compound as a carbon, nitrogen, sulfur, or potentially, a phosphorous source.

EXAMPLE 5: Directed Evolution of Enzymes

This metabolic selection method is also capable of facilitating the directed evolution of enzymes. One can use this 30 technique to screen known enzymes for mutations leading to higher efficiency, or to better specify optimal temperature or cofactor requirements, in the metabolic utilization of a compound. The mutations can be the result of natural evolution, the result of

PCR or chemical mutagenesis, or created through techniques like DNA shuffling.

EXAMPLE 6: Glucose to Ascorbic Acid Directly

Another permutation on this strategy that can be envisioned
5 is to find new pathways for already existing processes, e.g. selection for a new pathway for the conversion of glucose to ascorbic acid using only a few enzymatic steps. This is feasible using, for example, a strain for which the sequence of the entire genome is known, such as *E. coli* or *B. subtilis*. The genes for
10 the metabolism of glucose can be mutagenized such that the strain can no longer use glucose as a carbon/energy source, and then glucose-utilization pathways can be selected for as described in the previous examples.

EXAMPLE 7: Ascorbate Biosensor (Fig. 13)

As mentioned above, the *yiaJ* protein is thought to be a regulator for the *Yia* operon. The experiments of the invention indicate that the regulatory activity of *YiaJ* may be, in part, modulated by sensing ascorbate. Thus, it is currently believed that the "sensing" of ascorbate by *YiaJ* (perhaps through binding,
15 although the authors do not wish to be restricted to this interpretation) leads to the activation of the *Yia* operon, and thus the use of ascorbate as a carbon/energy source. This potentially results in an extremely sensitive "biosensor" for ascorbate. Thus, for example, it is envisioned that *yiaJ* could
20 be placed in a construct such that when *YiaJ* bound ascorbate a detectable signal resulted, i.e. instead of turning "ON" or "OFF" the *Yia* operon, *YiaJ* could turn "ON" or "OFF" a gene which produces a detectable signal, for example a gene for fluorescence (e.g. β -galactosidase), luminescence (e.g. luciferase), or color
25 (lac operon, or green fluorescent protein). Methods of constructing these signal constructs are well-known in the art (e.g. Simpson, et al. 1998. TIBTECH 16: 332-338; Applegate, et
30

a1. 1998. Applied Environ. Microbiol. 64: 2730-2735; Selifonova and Eaton, 1996. Applied Environ. Microbiol. 62: 778-783).

These biosensor constructs can also be used in the methods of the invention for screening for a metabolic selection pathway instead of using selection on an essential factor or element. In this case, the tester strain would be one that does not have the source to target pathway as determined by the absence of target being detected by the biosensor in the presence or the absence of the source compound. Thus, the biosensor would need to "sense" and to "react to" the presence of the target compound by any one of the methods described above. Following transfection of the library of nucleic acid from environmental sources, the resulting cells would be screened for the presence of the target compound using the biosensor. In order to facilitate the numbers of colonies that would need to be screened, this could be automated read in luminescent or fluorescent readers or sorted by FACS prior to further testing and identification of individual colonies. Although this requires more initial screening than selection using an essential element, this method offers an alternative approach when the appropriate tester strain or the metabolic pathway is not available for screening using an essential factor. Thus, the biosensor method provides the flexibility to identify pathways for compounds that are not metabolizable to an essential element, factor, or nutrient, but can be any compound for which a "biosensor" can be identified. Biosensors can be identified and created as described above.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those

skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the 5 invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains.

10 The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting 15 of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions 20 thereof, but it is recognized that various modifications are possible within the scope of the invention claimed.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in 25 terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

30 Other embodiments are within the following claims.

Claims

1. A method for screening for one or more nucleic acid sequences that express one or more products that convert a source compound into a target compound, comprising contacting a cell 5 with one or more test nucleic acid sequences, wherein said cell expresses one or more genes encoding one or more proteins that in the presence of said target compound provide a detectable signal, wherein said detectable signal indicates the presence of said one or more nucleic acid sequences.
- 10 2. The method of claim 1, wherein said one or more nucleic acid sequences encodes a metabolic pathway not normally present in said cell.
- 15 3. The method of claim 2, wherein said one or more nucleic acid sequences are selected from the group consisting of mutagenized DNA, environmental DNA, combinatorial libraries, and recombinant DNA.
- 20 4. The method of claim 3, wherein said environmental DNA is isolated from one or more sources selected from the group consisting of mud, soil, water, sewage, flood control channels, and sand.
5. The method of claim 3, wherein said mutagenized DNA is the result of enzyme mutagenesis wherein said mutagenesis is selected from the group consisting of random, chemical, PCR-based, and directed mutagenesis.
- 25 6. The method of claim 5, wherein said enzyme is selected from the group consisting of lactonases, esterhydrolases, and reductases.

7. The method of claim 1, wherein said detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

8. The method of claim 7, wherein said detectable signal
5 is growth.

9. The method of claim 1, wherein said target compound provides an element required for growth.

10. The method of claim 9, wherein said element is selected from the group consisting of carbon, nitrogen, sulfur, and phosphorous.
10

11. The method of claim 10, wherein said element is carbon.

12. The method of claim 9, wherein said target compound is selected from the group consisting of ascorbate and 2-KLG.

13. The method of claim 12, wherein said target compound is
15 ascorbate.

14. The method of claim 1, wherein said source compound is selected from the group consisting of 2-Keto-L-Gulonate, 2,5-Deoxy-Keto-Gulonate, L-Idonate, L-Gulonate, and glucose.

15. The method of claim 14, wherein said source compound is
20 2-Keto-L-Gulonate.

16. The method of claim 1, wherein said cell naturally expresses said one or more genes encoding said one or more proteins that in the presence of said target compound provide a detectable signal.

17. The method of claim 16, wherein said one or more proteins are one or more Yia operon-related polypeptides.

18. The method of claim 1, wherein said cell has been genetically manipulated to express said one or more genes 5 encoding one or more proteins that in the presence of said target compound provide a detectable signal.

19. The method of claim 18, wherein said one or more proteins are one or more Yia operon-related polypeptides.

20. The method of claim 18, wherein said one or more genes 10 encoding said one or more proteins are under the control of an inducible promoter.

21. The method of claim 20, wherein said inducible promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene.

15 22. The method of claim 1, wherein said cell grows on ascorbate and does not grow on 2-Keto-L-Gulonate.

23. The method of claim 22, wherein said cell is a bacteria.

24. The method of claim 23, wherein said bacteria is 20 *Klebsiella oxytoca*.

25. The method of claim 1, wherein said cell grows on 2-Keto-L-Gulonate and does not grow on 2,5-Deoxy-Keto-Gulonate.

26. An isolated, enriched, or purified nucleic acid molecule encoding one or more Yia operon-related polypeptides 25 selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

27. The nucleic acid molecule of claim 26, wherein said nucleic acid molecule comprises a nucleotide sequence that:

(a) encodes a polypeptide having the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12,

5 SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18;

(b) is the complement of the nucleotide sequence of (a); and

10 (c) hybridizes under highly stringent conditions to the nucleotide molecule of (a) and encodes a naturally occurring polypeptide.

28. The nucleic acid molecule of claim 26, further comprising a vector or promoter effective to initiate transcription in a host cell.

15 29. The nucleic acid molecule of claim 26, wherein said nucleic acid molecule is isolated, enriched, or purified from a bacteria.

30. The nucleic acid molecule of claim 29, wherein said bacteria is *Klebsiella oxytoca*.

20 31. A nucleic acid probe for the detection of nucleic acid encoding one or more Yia operon-related polypeptides, selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS, in a sample.

25 32. The probe of claim 31, wherein said polypeptide is a fragment of the protein encoded by the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18.

33. A recombinant cell comprising a nucleic acid molecule encoding one or more Yia operon-related polypeptides selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

5 34. The cell of claim 33, wherein said polypeptide is a fragment of the protein encoded by the amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18.

10 35. An isolated, enriched, or purified Yia operon-related polypeptide selected from the group consisting of YiaJ, YiaK, YiaL, ORF1, YiaX2, LyxK, YiaQ, YiaR, and YiaS.

15 36. The polypeptide of claim 35, wherein said polypeptide is a fragment of the protein encoded by the full length amino acid sequence set forth in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18.

37. The polypeptide of claim 35, wherein said polypeptide is isolated, enriched, or purified from bacteria.

20 38. The nucleic acid molecule of claim 37, wherein said bacteria is *Klebsiella oxytoca*.

39. An isolated, enriched, or purified nucleic acid molecule, wherein said nucleic acid molecule comprises the nucleotide sequence set forth in SEQ ID NO:19.

25 40. The nucleic acid molecule of claim 39, wherein said nucleic acid molecule comprises:

(a) one or more nucleotide sequences that are set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9;

(b) the complement of the nucleotide sequence of (a);

5 (c) nucleic acid that hybridizes under stringent conditions to the nucleotide molecule of (a);

(d) the full length sequence of SEQ ID NO:19, except that it lacks one or more of the sequences set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6,

10 SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9; and

(e) the complement of the nucleotide sequence of (d).

41. The nucleic acid molecule of either of claims 39 or 40, further comprising a vector or promoter effective to initiate transcription in a host cell.

15 42. The nucleic acid molecule of claim 41, wherein said vector or promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene.

43. The nucleic acid molecule of claim 39, wherein said nucleic acid molecule is isolated, enriched, or purified from a
20 bacteria.

44. The nucleic acid molecule of claim 43, wherein said bacteria is *Klebsiella oxytoca*.

45. A recombinant cell, comprising the nucleic acid molecule of claim 42.

25 46. A recombinant cell useful for screening for one or more nucleic acid sequences that express one or more products that convert a source compound into a target compound, wherein said cell expresses one or more genes comprising an inducible promoter, and wherein said one or more genes encodes one or more

proteins that in the presence of said target compound and an inducer provide a detectable signal, wherein said detectable signal indicates the presence of said one or more nucleic acid sequences.

5 47. The recombinant cell of claim 46, wherein said one or more nucleic acid sequences encodes a metabolic pathway not normally present in said cell.

10 48. The recombinant cell of claim 47, wherein said one or more nucleic acid sequences are selected from the group consisting of mutagenized DNA, environmental DNA, combinatorial libraries, and recombinant DNA.

15 49. The recombinant cell of claim 48, wherein said environmental DNA is isolated from one or more sources selected from the group consisting of mud, soil, water, sewage, flood control channels, and sand.

50. The recombinant cell of claim 48, wherein said mutagenized DNA is the result of enzyme mutagenesis wherein said mutagenesis is selected from the group consisting of random, chemical, PCR-based, and directed mutagenesis.

20 51. The method of claim 50, wherein said enzyme is selected from the group consisting of lactonases, esterhydrolases, and reductases.

25 52. The recombinant cell of claim 46, wherein said detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

53. The recombinant cell of claim 46, wherein said detectable signal is growth.

54. The recombinant cell of claim 53, wherein said cell requires the presence of said target compound and said inducer for growth.

55. The recombinant cell of claim 54, wherein said target compound is selected from the group consisting of ascorbate and 2-Keto-L-Gulonate.

56. The recombinant cell of claim 46, wherein said one or more genes are under the control of said inducible promoter.

10 57. The recombinant cell of claim 56, wherein said inducible promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene.

58. The recombinant cell of claim 56, wherein said one or more proteins comprise one or more Yia operon-related polypeptides.

15 59. The recombinant cell of claim 58, wherein said cell naturally expresses said one or more genes.

60. The recombinant cell of claim 58, wherein said cell has been genetically manipulated to express said one or more genes.

20 61. The recombinant cell of claim 58, wherein said cell is a bacteria.

62. The recombinant cell of claim 61, wherein said bacteria is *Klebsiella oxytoca*.

25 63. A method for identifying a substance that modulates the conversion of a source compound to a target compound, comprising: contacting a cell with nucleic acid, wherein said nucleic acid expresses a product that converts a source compound into a

target compound, and wherein said cell expresses one or more proteins which in the presence of said target compound provide a detectable signal;

5 contacting said cell with a test substance; and
monitoring said detectable signal, wherein said detectable signal indicates the presence of said substance.

64. The method of claim 63, wherein the substance is selected from the group consisting of antibodies, small organic molecules, peptidomimetics, and natural products.

10 65. The method of claim 64, wherein said detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

15 66. The method of claim 65, wherein said detectable signal is growth, and wherein said target compound is metabolizable to an element selected from the group consisting of carbon, nitrogen, sulfur, and phosphorous.

67. The method of claim 66, wherein said element is carbon.

68. The method of claim 63, wherein said source compound is selected from the group consisting of 2-Keto-L-Gulonate, 2,5-
20 Deoxy-Keto-Gulonate, L-Idonate, L-Gulonate, and glucose.

69. The method of claim 63, wherein said one or more proteins are one or more Yia operon-related polypeptides.

70. The method of claim 69, wherein said Yia operon further comprises a vector or promoter effective to initiate
25 transcription in a host cell.

71. The method of claim 70, wherein said vector or promoter comprises the *trp-lac* hybrid promoter, the *lacO* operator, and the *lacI^q* repressor gene.

72. A method for detecting the presence, absence, or amount
5 of a compound in a sample comprising:

contacting said sample with a cell, wherein said cell expresses one or more genes encoding one or more proteins that in the presence of said compound provide a detectable signal that indicates the presence, absence, or amount of said compound.

10 73. The method of claim 72, wherein said compound is ascorbate.

74. The method of claim 72, wherein said detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

15 75. The method of claim 72, wherein said one or more genes comprises *yiaJ*.

76. The method of claim 75, wherein said one or more genes further comprises a promoter transcriptionally linked to a reporter gene.

20 77. The method of claim 76, wherein *YiaJ* is naturally expressed in said cell.

78. The method of claim 76, wherein said cell has been genetically manipulated to express said *yiaJ*.

25 79. The method of claim 76, wherein the expression of said reporter gene is regulated by the binding of *YiaJ* to said promoter.

80. The method of claim 72, wherein said cell is a bacteria.

81. The method of claim 80, wherein said bacteria is *Klebsiella oxytoca*.

5 82. An isolated, purified, or enriched nucleic acid molecule encoding YiaJ and a reporter gene.

83. The nucleic acid molecule of claim 82, further comprising a promoter transcriptionally linked to said reporter gene.

10 84. The nucleic acid molecule of claim 83, wherein the expression of said reporter gene is regulated by the binding of YiaJ to said promoter.

15 85. A recombinant cell for detecting the presence, absence, or amount of a compound in a sample comprising the nucleic acid molecule of either of claims 82 or 83.

86. A recombinant cell for detecting the presence, absence, or amount of a compound in a sample, wherein said cell expresses one or more genes encoding one or more proteins that in the presence of said compound provide a detectable signal, wherein said signal indicates the presence, absence, or amount of said compound.

87. The recombinant cell of claim 86, wherein said detectable signal is selected from a group consisting of growth, fluorescence, luminescence, and color.

25 88. The recombinant cell of claim 86, wherein said one or more genes comprises *yiaJ*.

89. The recombinant cell of claim 88, wherein said one or more genes further comprises a promoter transcriptionally linked to a reporter gene.

90. The recombinant cell of claim 89, wherein YiaJ is naturally expressed in said cell.
5

91. The recombinant cell of claim 89, wherein said cell has been genetically manipulated to express said *yiaJ*.

92. The recombinant cell of claim 89, wherein the expression of said reporter gene is regulated by the binding of
10 YiaJ to said promoter.

93. The recombinant cell of claim 86, wherein said cell is a bacteria.

94. The recombinant cell of claim 93, wherein said bacteria is *Klebsiella oxytoca*.

15 95. A method of selection for a nucleic acid sequence encoding a metabolic pathway from a source compound to a target compound comprising:

(1) identifying an organism that metabolizes a target compound to provide an essential element;

20 (2) identifying one or more genes responsible for the metabolism of said target compound to said essential element;

(3) expressing said one or more genes under the control of an inducible promoter, whereby said target compound is metabolized in the presence of an inducer and not in the absence
25 of said inducer;

(4) expressing nucleic acid sequences potentially encoding said metabolic pathway in said recipient organism; and

(5) selecting said recipient organism for growth on said source compound in the absence of said target compound and in the

presence of said inducer, wherein growth on said source compound in the absence of said target compound and in the presence of said inducer indicates the presence of said nucleic acid sequence.

5 96. The method of claim 95, wherein said essential element is selected from the group consisting of carbon, phosphorous, nitrogen, and sulfur.

97. The method of claim 96, wherein said essential element is carbon.

10 98. The method of claim 95, further comprising the transfer of said one or more genes to a highly genetically manipulatable recipient organism, such that said recipient organism metabolizes said target compound to provide an essential element.

Figure 1

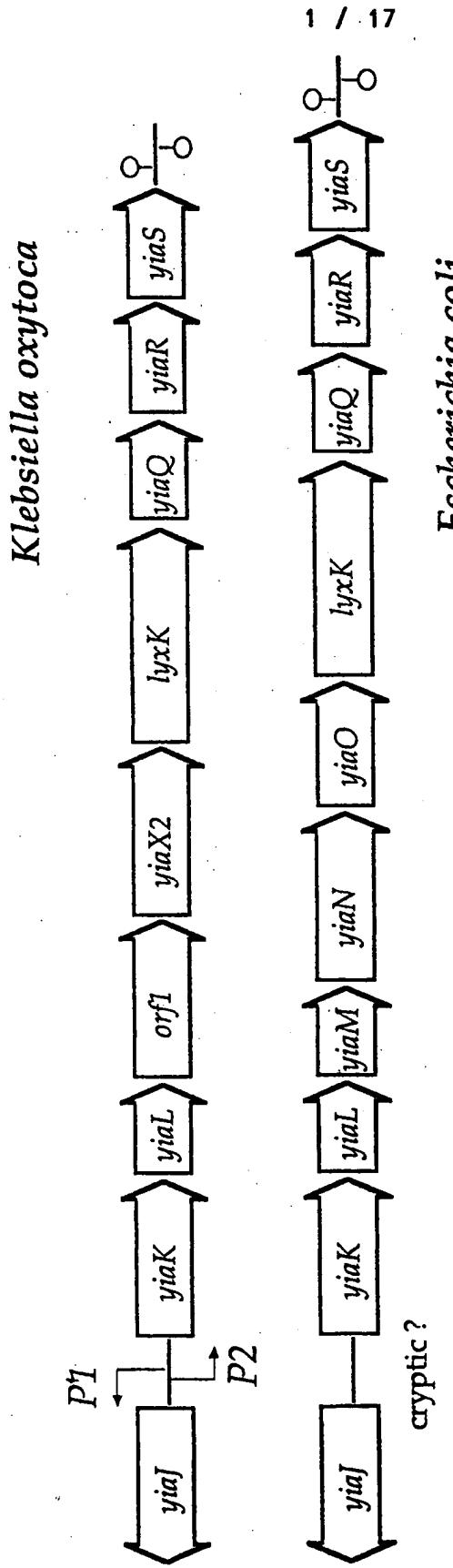


Figure 2A

GGATCCGGGGCGAAAGCGGAGACGCCAGAACAGTCCTGGTCTGCTGATGGGACACCACGCAGGCACCTCACAGGT
 ACAGCAGCCGATCTCCGCATCCGCAGAATAAACCGATTCACTCTCTCATGGGGATAAAAACGCAGAGTGC
 CAGAAAAAACCGCTTCCCTCTCCCTTGATCTGAATGGAGTCAGCGGGTTCTCTCAGATGTCCGGATTATCTGG
 80
 160
 240

* G E R V S F G L E R S I A E A T D R L P K L
 TCATTTGCCCTAACCTCCGCACGGAAAAGCCCAGTCGCAGAAAATGCCCTGCCATGCCGTAGCGGCTTGAGTA
 320

L N K E G V Q K L R S T S L S I S I A Y P V R G H I D
 AATTTTCTCTCCACCTGCTGAGGCGCAGTGTGATAGAGAGATAGAAATGCCATAAGGCACGCCATGGATATCA
 400

F V P V A L C S V G L E N E E R D M A M N R E R I Q A
 AAAACGGGGACAGCCAGGCACGCCAGCTCGTCTCCCTGTCATGCCATATTCGCTCGGGATCTGCC
 480

L E D H M A P L G T I T N R T L P Q I I E O H S N W
 CAGTTCATCATGCATCGCAGGCARGCCGGTAATGGTATTACGGGTAGCGGCTGGATAATCTCTGGTGTGAATTCCAGT
 560

Y S E V Y D P H G F A M Y I K A W Q R A T C R M H Q G
 AGCTCTCAACGTAGTCAGGATGCCAAACGCCATATAATCTTGCCCATGGCCAGCAGTACAGCGCATGTGGCCA
 640

I Y A R T R L M G T T P E L K Y I L I A H D D E R S S
 ATATAGGCCGCGTACGCAGCATACCGGTGGTGGCTCCAGCTTATAATCAGGATCGCTGGTATCTCACGGCTGGA
 720

F N V T E G T A L N L A E L H P A A V H I I N L S S
 GAAGTTCACCGTCTCGCCGGTGGCAGGTTAAgCGCCTCAAGATGCGGCCGCCGACGTGGATAATATTCAAGCAGACA
 800

L A K Q G V R I F K T T L A Y S G A A P A P T V Y G C
 ACGCCCTTGGCCAACCGGATAAATTTCGTCAGCGCATAGCTCCCGCCGGGGcAGCGTCACGTACCCGCA
 880

S Q L G Q L L R H V T S K N L G A L E S L H A V P C G
 GACTGCAGCCCTGTAATAAGCGATGAACGGTACTTTGTCAGTCCCGCCAGTTCCGACAGATGCGCACGGGACAGCC
 960

N P Y N S L I E I L M L G R F L S Q S G A P R E K D
 ATTGGATAATTACTCAGGATCTCAATTAGCATCAACCCACGAAAAAGGCTCTGACTTCCGGCAGGCCCTCTCTTATCTT
 1040

Q T N E S E K T G M
 GCGTGTCTCGCTTCTTGTGCCATCGCTCCGCTCCATTGGCTGAGATGGTAGCGCAAAGTGTGTTTC
 1120

yiaj ←
 AGTCACGATCTGAACCGAAAAAACACAACCTTTATGATTTTATGATTTTAAAATAACGCTGCCGTTGATCTGACAA
 AAATTGATCGCTATATTGAAATCAGATTGCAAGTGGATAAAAAGCGATCAACTCTGACCAGGAAA
 → *yiaK*
 CAGCAATGAAAGTCACGTTGAGCAGTAAAGAGGCATTCAATCGGTACTGCTGGAGcgcgtcgccggGGAAACC
 M K V T F E Q L K E A F N R V L L D A C V A R E T
 1360

GCCGATGCCCTGCCAGAAATGTTGCCGCACCACCGAATCCGGCTCTATTCTCACGGCTGAACCGCTTCCCTGCTT
 A D A C A E M F A R T T E S G V Y S H G V N R F P R F
 1440

CATCCAGCAGTTGGATAACGGCGACATTATCCCTGAGGCTCAACCCAGCGGGTGACCACGCTCGGCCATCGAACAGT
 I Q Q L D N G D I I P E A Q P Q R V T T L G A I E Q W
 1520

GGGATGCTCAGCGTCCATGGCAACCTGACGGGAAAAAGATGGATGGCTGGCCATTGAGCTGGCTCCGATCACGGT
 D A Q R S I G N L T A K K M M D R A I E L A S D H G
 1600

ATCGGCCCTGGTGCCTTACGTAATGCTAACCAACTGGATGCGCGGCCAgcTACGGCTGGCAGGGCGGGAAAAAGGCTA
 I G L V A L R N A N H W M R G G S Y G W Q A A E K G Y
 1680

CATCGGTATCTGCTGGACCAACTCCATGCCGTTATGgcGCCATGGGGCGTAAAGAGTGGCTATCGGTACCAACCCGC
 I G I C W T N S I A V M A P W G A K E C R I G T N P L
 1760

TGATCGTGCCTTCCGCTGACGGCGATCACCATGGTGGATATGTCGATGTCGATGTTCTCTACGGCATGCTGGAGGTT
 I V A I P S T P I T M V D M S M S M F S Y G M L E V
 1840

Figure 2B

AACCGCCTGCCGGCCGAACTGCCGTGGACGGCGATTGACGATGACGGTGTGACCAAAGAGCCGGGACGAT	1920
N R L A G R E L P V D G G F D D D G R L T K E P G T I	
CGAGAAAAAATGCCGCATTTACCCATGGGCTACTGGAAAGGTTCCGGCTGTCATCGTGTGGATATGATTGCCACCC	2000
E K N R R I L P M G Y W K G S G L S I V L D M I A T L	
TCCCTCTCCAACGGATCGTCGGTTGCCGAAGTGACCCAGGAAAACAGCGATGAATATGGCCTTCGAGATCTTCATCGCT	2080
L S N G S S V A E V T Q E N S D E Y G V S Q I F I A	
ATTGAAGTGGATAAGCTGATCGACGGCGAACCCGGACGCCAACGGATTATGGATTTCATCACCAACGCCGA	2160
I E V D K L I D G A T R D A K L Q R I M D F I T T A E	
GC CGCCGATGAAAATGTGGCGTCCGCTTCCCTGGCATGAATTACCCGCTGCTGGATGAAAACGCCGAACGCCA	2240
R A D E N V A V R L P G H E F T R L L D E N R R N G I	
TTACCGTCGATGACAGCGTATGGCCAAAATTCAAGCGCTGTAAGGAGCTACCCATGACAGCGTATGGCCAAAATTCA	2320
T V D D S V W A K I Q A L *	
→ <i>yial</i>	
GGCGCTGTAAGGAGCTACCCATGATTTGGTCAATTGCTCACCTAACCTAACCGTGTGCTGCCCGGCCATTGAGCG	2400
M I F G H I A Q P N P C R L P A A I E R	
GGCGCTTGTATTCCTGCCACGACGGATTCCACGGCTGGCACCCGGCTGTTGGAAATCGACGGCAAACATCTTCG	2480
A L D F L R T T D F H A L A P G V V E I D G Q N I F A	
CGCAGGTTATCGACTTAACCACTCGCGATGCCGCTGAAAATCGTCGGAGGTCCACCGTCGCTATCTGGATATCCAGTT	2560
Q V I D L T T R D A A E N R P E V H R R Y L D I Q F	
CTGGCATCGGGCGAAGAAAAATCGGTATGCCATTGATACCGGAATAATCAAATCAGCAATCTTATTAGAACAGCG	2640
L A S G E E K I G I A I D T G N N Q I S E S L L E Q R	
CGATATTATTTTATCACGACAGCGAACATGAATCGTCTTGGAAATGACGCCAGGCAACTATGCGATATTTCCCGC	2720
D I I F Y H D S E H E S F F E M T P G N Y A I F F P Q	
AAGATGTTATCGTCCTGGATGTAATAAAACTGTAGCCACGCCATCCGAAATAGCTGTTAAAGTCGCTATTCAGTT	2800
D V H R P G C N K T V A T P I R X I V V R V A I S V	
→ <i>orf1</i>	
TTATAAGAAGGAGCACAAATGAATTGAAATAACCGTTACATTATCGGTGCGTACCCCTGTGCCCCCTGTGCACCC	2880
L *	
M N S N N T G Y I I G A Y P C A P C A P S	
CATTTCACCAAAAGAGTCGAAGAGGAAGAGAGTGGAATTCTGGCGAGCTCTCGACACCCGGATATCGCGGCTGGAG	2960
F H Q K S E E E M E F W R Q L S D T P D I R G L E	
CAACCTGCCCTACCCCTGCCCTGAACATCTCATCCGCTGGCGACGAGTGGTTATTGCCCATACCCGGACACTGGCA	3040
Q P C L P C L E H L H P L G D E W L L R H T P G H W Q	
GATTGTCGTTACGCCATCATGGAAACCATGCGCCGCCGGTGAAACGGCGCTTGGCTGGCTCCAGCGACGAAA	3120
I V V T A I M E T M R R R G E N G G F G L A S S D E T	
CGCAGCGCAAAGCTGGTGGAGTACTATGCCACCTGCAGCAGAAGATCGTAAATCAATGCCAATACCGCCGGAAAG	3200
Q R K A C V E Y Y R H L Q Q K I A K I N G N T A G K	
GTCATTGCCCTTGAGCTCACGCCGCCGCTGGCGGGCAATGCCAACGTGGCTCAGGCTACCGACGCCCTTGCCCGTTC	3280
V I A L E L H A A P L A G N A N V A Q A T D A F A R S	
ATTAAAAGAAATTACCCGCTGGGACTGGTCTGCGAGCTGGCTGGAGCACTGCGACGCCATGCCAGGGCAGCGCCGC	3360
L K E I T R W D W S C E L V L E H C D A M T G S A P R	
GCAAAGGATTTGCCGTTAGAAAACGTGCTGGAAGGCCATTGCCGATTATGACGTTgGCATTGTATTAACTGGCGCGT	3440
K G F L P L E N V L E A T A D Y D V G I C I N W A R	

Figure 2C

TCGGCCATTGAAGGGCGGAATACCGTCTACCGCTACCCATAACGAGGTAAACGGCAGGAAGCTCGGCGCT 3520
 S A I E G R N T V L P L T H T Q Q V K R A G K L G A L

 GATgTTTCTGGCACGACCGAgACCGCGAGTACCGCGAATGGCAGGATTACACCGCCGTCGCGCTTCTGCCGC 3600
 M F S G T T Q T G E Y G E W Q D L H A P F A P F C P Q

 AgAGCCTGATGACCACCGAACACGCTCGTAATTACCGCTGCAGGAACGCCCCCTGCAATTTCAGGCATTAAA 3680
 S L M T T E H A R E L F A C A G T A P L Q F S G I K

 TTACTGAAATTAAATGCCAGCGCAAACGTTGATCATCGCATCGCATATTACCGGACGGCATCTCGCGCTAAACAAGC 3760
 L L E I N A S A N V D H R I A I L R D G I S A L K Q A
 → *yiaX2*
 ACAATAATAATAATCACCTTCATCACCGAGAATTTAAATATTACGAGACTATAAGATgATAACCTCTAACTCTA 3840
 Q * M N I T S N S T

 CAACCAAAGATATAACCGGCCAGCGCTGGTTAAGAATCATTCCGCTATACTGATCACTTGATTTCTTATATGGAC 3920
 T K D I P R Q R W L R I I P P I L I T C I I S Y M D

 CGGGTCAATATTGCCCTTGCGATGCCGGAGGTATGGATGCCACTTAGGTATTCCGCCACCATGGCGGGCTGGCGG 4000
 R V N I A F A M P G G M D A D L G I S A T M A G L A G

 CGGTATTCTTCTTATCGGTTATCTTTACAGGTTCCCGGGAAAATTGCCGTTACCGTAGCGGTAAGAAATTAA 4080
 G I F F I G Y L F L Q V P G G K I A V H G S G K K F I

 TCGGCTGGTCGCTGGCGCTGGCGGTCACTCCGTGCTGACGGGTTAATTACCAATCAGTACCAAGCTGCTGGCCCTG 4160
 G W S L V A W A V I S V L T G L I T N Q Y Q L L A L

 CGCTTCTTACTGGCGTGGCGGAAGGCGGTATGTCGCCGTCGTTCTCACGATGATCAGTAACCTGGTCCCCGACGCTG 4240
 R F L L G V A E G G M L P V V L T M I S N W F P D A E

 ACGCGGTCGCCAACCGCATTGTCATTATGTTGCGGATTACCGCCCCACTCTCAGGCTGGATT 4320
 R G R A N A I V I M F V P I A G I I T A P L S G W I I

 TCACGGTTCTCGACTGGCGCTGGCTTTATTATCGAAGGTTGCTCTCGCTGGTTCTGGTTCTGTGGCATAACACC 4400
 T V L D W R W L F I I E G L L S L V V L V L W A Y T

 ATCTATGACCGTCCGCCAGGAAGCGCGCTGGATTTCGAAGCAGAGAGCGCTATCTGGTCGAGACGCTGGCGGGAGCA 4480
 I Y D R P Q E A R W I S E A E K R Y L V E T L A A E Q

 AAAAGCCATTGCCGGACCGAGGTGAAAAACGCCCTCTCTGAGCGCCGTTCTCCGACAAAACCATGTGGCAGCTTATCG 4560
 K A I A G T E V K N A S L S A V L S D K T M W Q L I A

 CCCTGAACCTCTTCTACCAgACCGCATTACGGCTACCCCTGTGGCTACCCACCATTCTGAAAGAATTGACCCATAGC 4640
 L N F F Y Q T G I Y G Y T L W L P T I L K E L T H S

 AGCATGGGCAGGTCGGCATGCTTGCCATTCTGCCGTACGTGGCGCCATTGCTGGGATGTTCTGTGTTCTCCCTTC 4720
 S M G Q V G M L A I L P Y V G A I A G M F L F S S L S

 AGACCGAACCGTAAACGCAAGCTTCGCTGCCGCTGGCTGGGATTCTCCCTGCAATCGCCGCTGGCTGGCG 4800
 D R T G K R K L F V C L P L I G F A L C M F L S V A L

 TgAAAAACCAAATTGGCTCTCTATGCCGCTGGCTGGCTGGGATTCTCCCTGCAATCGCCGCTGGCTGGCG 4880
 K N Q I W L S Y A A L V G C G F F L Q S A A G V F W

 ACCATCCGGCACGCTGTTAGCGCGGAATGGCGGGCGCCGCGCGGGTTATCACGCCCTGGCAACCTCGGGG 4960
 T I P A R L F S A E M A G G A R G V I N A L G N L G G

 ATTTGTGGCCCTATGCCGCTGGGTGCTGATCACGTTgTACAGCAAAGACGCTgGCCGCTATTGCCCTGGCGATCTCCC 5040
 F C G P Y A V G V L I T L Y S K D A G V Y C L A I S L

Figure 2D

TGGCGCTGGCCGCGCTGATgGCGCTgCTGCTGCCGGCAAATGCGATGCCgGTGCTGCGCCGGTaAAgACgATAAaTCCA 5120
 A L A A L M A L L L P A K C D A G A A P V K T I N P

CATAAACGCACTGCGTAAACTCGAGCCCCGGCGCCTgCGCCTGCCGGCCTGCCAAATATGCCGGTTACCCGGTaAC 5200
 H K R T A *

→ *lyxK*

AATgAGATGCgAAAqATGAGCAAqAAACAgGCCTCTGGCTGGGTATTGATTGCGGCGGCACCTATCTGAAGCCGGTT 5280
 M S K K Q A F W L G I D C G G T Y L K A G L

ATATGACGCCGAAGGTATGAACATGGCATGGCTGGCAAGCGCTACCGACGATGTCGCCCTGCCGGTTACGCCGAAC 5360
 Y D A E G H E H G I V R Q A L R T M S P L P G Y A E R

GCGACATGCCAGCTGGCAACACTGCGGGGACCATTCGCCGGTATTACAGCAGGCAGGTATCCGGCAACAG 5440
 D M R Q L W Q H C A A T I A G L L Q Q A G V S G E Q

ATTAAGGCCTGGGCATCTCCGCTAGGGCTAAGGGCTTTCTCCCTCGATAAGCAGGATCCGGCTGGGTAAACGCCAT 5520
 I K G V G I S A Q G Q G L F L L D K Q D R P L G N A I

CCCTCTCCGATCGTCGGCGCTGAAAATCGTCAGCGCTGGCAGCGGGACCGTATTCCCGAACGGCTATCCCGTTA 5600
 L S S D R R A L K I V Q R W Q R D R I P E R L Y P V T

CCCGCCAGACGCTGTGGACCGGACATCCGGCTTGTGGCTGGTAAAGAGAATGAACCCAGCGCTACGGCAA 5680
 R Q T L W T G H P A S L L R W V K E N E P Q R Y A Q

ATTGGCTCGTGTGATGGGCATGACTATCTGGCTGGCTTAACCGCGCGAAGGGCTGCGAGGAGAGCAACATCTC 5760
 I G C V M M G H D Y L R W C L T G A K G C E E S N I S

CGAGTCCAACCTCTACAACATGGCATGGCCAGTACGACCCGCGCTGACCGAGTGGCTGGCATGGTAAATCGATA 5840
 E S N L Y N M A M G Q Y D P R L T E W L G I G E I D S

GCGCGCTGCCCGGTTGTAGGGCTAGCCGAAATTGCGGGGAGATCACCGCTCAGGCAGCCGCTTAAACGGCTGGCG 5920
 A L P P V V G S A E I C G E I T A Q A A A L T G L A

GCGGGTACTCCGTCGTTGGCGGCCGTTTGACGTGGTCTCCACCGCCCTTGCGCCGGATTGAGGATGAGTCGACCC 6000
 A G T P V V G G L F D V V S T A L C A G I E D E S T L

CAATGCGGTGATGGGACCTGGCGCTCACTAGCGGTATCGCTCACGGCTGGCGACCATGAGGCCACCCCTAACGTCT 6080
 N A V M G T W A V T S G I A H G L R D H E A H P Y V Y

ATGGCCGCTACGTCAATGACGGCACTATCGTCACGAAGCCAGCCGACCTCATCCGCAACCTGAATGGTTTACC 6160
 G R Y V N D G Q Y I V H E A S P T S S G N L E W F T

GCCCAGTGGGCGATCTCGTTGATGAGATCAATCAGGCCGCGCCAGCTGCCAAAGCCGGAGCGAGCTGTTTT 6240
 A Q W G D L S F D E I N Q A V A S L P K A G S E L F F

TCTGCCGTTCTGTATGGCAGCAACGCCGGCTGGAGATGACCTGCCCTTACGGCATGCGAGGCCGCTGCATACCGCG 6320
 L P F L Y G S N A G L E M T C G F Y G M Q A L H T R A

CGCACCTGCTGCAGGCCGGTTATGAAGCGTGGTATTAGCCATATGACCCACCTCAGCGTATGCGCAACGCTTACA 6400
 H L L Q A V Y E G V V F S H M T H L S R M R E R F T

AACGTTCAAGGCCGCGCTCACCGGCCGCCGCCGCACTCCGACGCTGGATGCGAGATGCTGGCGACGTAAGCGCTT 6480
 N V Q A L R V T G G P A H S D V W M Q M L A D V S G L

ACGCATTGAACCTCCGAAGGTGGAAGAGACCGGCTGGCTGGCCACCTGCCGCTCGGTGAGATGCTGGCGACGTAAGCGCTT 6560
 R I E L P K V E E T G C F G A A L A A R V G T G V Y R

GCAGcTTAGCGAAGGCCGGCGCCGCCGGCAGCACCCGGTGCACGcTGCTGCCGATATGACCGCCACGCCGCTAT 6640
 S F S E A R R A R Q H P V R T L L P D M T A H A R Y

Figure 2E

→ *yiaQ*

cAGCGCAAATACGCCACTACtTGCAATTGATTGAAGCACTACAGGGTATCAGCCGTATAAGGAGCACGCATTATG 6720
 Q R K Y R H Y L H L I E A L Q G Y H A R I K E H A L * M

AGCCGACCATTACTGCAGCTGGCGcTCGACCATAccAGCCTTCAGGCTGCGCAGCGCATGTCGCCCTGCTACAGGATCA 6800
 S R P L L Q L A L D H T S L Q A A Q R D V A L L Q D H

CGTTGATATTGTGGAGGCCGGAACCATCCTCTGCTTAACCGAAGGGCTTACGCCGGTTAACGCCCTGCCGCCAGTGTc 6880
 V D I V E A G T I L C L T E G L S A V K A L R A Q C P

CGGGGAAGATCATCGCCGACTGAAAGTCGCCGACGCCGGTGAAACCCCTGGCGCAGCAGGCCCTTGGCCTGGCGCC 6960
 G K I I V A D W K V A D A G E T L A Q Q A F G A G A

AACTGGATGACCATATTGCGCCGACCGCTGCCACGGTCGAGAAAGGCCACGCCGTGGCCAGGCCCTGCCGGTGA 7040
 N W M T I I C A A P L A T V E K G H A V A Q A C G G E

AATTCAAGATGGAGCTTCGGCAACTGGACGCTGGATGACGCCCGCGCTGGTACCGTACGGCGTCCATCAGGCGATT 7120
 I Q M E L F G N W T L D D A R A W Y R T G V H Q A I Y

ACCATCGCGGACCGATGCCAGGCCAGCGGGCAGCAGTGGGGGAGGCCAGTGGCGCGCATGAAAGCGCTGTCCGAT 7200
 H R G R D A Q A S G Q Q W G E A D L A R M K A L S D

ATTGGCCTTGAGCTATCGATTACCGCGGATTACCCCAGCGATCTACCGCTGTTCAAAGATATCAACGTCAAAGCCTT 7280
 I G L E L S I T G G I T P A D L P L F K D I N V K A F

TATTGCCGGCGCGCGCTGGCAGGCCGCCATCCGGCGGGTTGCCGCCATTCCACCGCAAATCGACGCTATCT 7360
 I A G R A L A G A A H P A R V A A E F H A Q I D A I W

→ *yiaR*

GGGGAGAACAGCATGGTAACCACCGTTAGGTATTATGAAAAAGCGCTGGCAAGGATCTAGCTGGCTGAGCGCT 7440
 G E Q H A * M R N H P L G I Y E K A L A K D L S W P E R L

GGTACTGGCCAAAAGCTGGTTTGATTTGCGAAATGTCGGTGGACGAGACCGATGAACGCCCTTCGCCCTGGAGT 7520
 V L A K S C G F D F V E M S V D E T D E R L S R L E W

GGACCCCGGCCAGCGCGCATCGCTGGTGAGCGCGATGCTGGAAACCGCGCTGCCATTCCCTCGATGTGCTTGTCCGCC 7600
 T P A Q R A S L V S A M L E T A V A I P S M C L S A

CATCGCCGTTCCCCCTGGCAGCCCGATGAAGCGGTACCGATCGGCCGAGAGATTATGACCAAAGCCATCGCCT 7680
 H R R F P F G S R D E A V R D R A R E I M T K A I R L

GGCGCCGATCTGGGATCCGCACCATCCAGCTGGCGGGTACCGCTATTACGAAGAGCATGATGAAGGCACCCGGC 7760
 A R D L G I R T I Q L A G Y D V Y Y E E H D E G T R Q

AGCGTTTGCGAAGGGCTGGCTGGCGTAGAACAGGCCGCCGCAGGTAATGCTGGCGTGGAGATCATGGAC 7840
 R F A E G L A W A V E Q A A A A Q V M L A V E I M D

ACCGCCTTATGAACCTCATCGCAAATGGAAAAGTGGGACGAGATGCTTCGTCACCGTGGTTACCGTCTACCGGA 7920
 T A F M N S I S K W K K W D E M L S S P W F T V Y P D

CGTCGGCAACCTCAGGCCCTGGGAAACGACGTACCGCCGAGCTGAAGCTGGCATCGATGTATGCCGCCATCCACC 8000
 V G N L S A W G N D V T A E L K L G I D R I A A I H L

TGAAAGATAACGCTGCCGTGACCGACGATAGCCCTGGCAGTCCGGACGTGCGTTCGCGAAGGATGCGTCGATTTT 8080
 R D T L P V T D D S P G Q F R D V P F G E G C V D F

GTCGGCATTTAAAGACGcTGCGGAGCTGAACCTACCGCGGTCACTTTGATTGAGATGTCGGACGGAGAAAGCCAGCGA 8160
 V G I F K T L R E L N Y R G S F L I E M W T E K A S E

→ *yiaS*

Figure 2F.

GCCGGTGTGGAGATTATCCAGGCCCGCCTGGATCGAATCACGGATGCCAGGAAGGGGATTACATGTTAGAACACT P V L E I I Q A R R W I E S R M Q E G G F T C *	8240
M L E Q L	
GAAAGCCGAGGTACTGGCCGAAACCTGGCCCTCCCCGACACGGCCTGGTCACCTTACCTGGGCAACGTCAGGGGG K A E V L A A N L A L P A H G L V T F T W G N V S A V	8320
TCGATGAAACCGCGAAGCTGATGGTCATTAAGCtTCCGGCGTCAAAATGAGGTGATGACCGCCGACGATATGGTGGTC D E T R K L M V I K P S G V E Y E V M T A D D M V V	8400
GTAGAGATGCCAGCGTAAAGCTGTTGAAGCGGTAAAAACCCCTTCAGATAACGCCAACGCATCTGGCGCTTATCG V E M A S G K V V E G G K K P S S D T P T H L A L Y R	8480
CCGCTATCCGAGATCGCGGGATCGCATAACCCACTCCGCCACGCCAGCATCTGGCGCAGGCCGGCTCGATCTCC R Y P Q I G G I V H T H S R H A T I W S Q A G L D L P	8560
CcGCCTGGGCACCACCCACGCCACTACTTCTATGGCGCGATCCCCTGTACCCGACGGATGACCGTTGAGGAGATAAC A W G T T H A D Y F Y G A I P C T R R R M T V E E I N	8640
GGCGAGTATGAGTATCAGACCGCGAGGTGATTATCAAAACCTTGAAACAGCGCCGCTGGATCCGGCAAATCCGGC G E Y E Y Q T G E V I I K T F E Q R G L D P A Q I P A	8720
GGTATTGGTCCATTACACGGCCCTTGCCTGGGTAAAGACGCCGCCAGCGCTACATAACGCCGTGGTGGCTGG V L V H S H G P F A W G K D A A D A V H N A V V L E E	8800
AGTGCCTACATGGCCTCTCTCGGCCAGTGGCACAGCTCCGGATATCGAGTCTGAACGCTCGATAAACACTAT C A Y M G L F S R Q W P Q L P D M Q S E L L D K H Y L	8880
CTCGTAAACACGGCGAACGCTATTACGGGAAACTAGTCCCGGAACTCCCGGATAAGCGCTTGGCCCCGG R K H G A N A I T G K T S P A E L P G	8960
GGGAAGCGTGCAGGATGTTGCTGAACCTTCCGGAGCGATGCTGCGCATCTGTCGGCTACCGTCCCCGGCTCTGC 9040	
GGTCAGCACCGCGCCGGAAAACCCATCAACCTACGCCGAATTAAATATGCTCTGGAGTAACGACGCTTCCACGCC GCCGGTCCAGGCTGGTGTGCTGGAAAATCTGCGAAAATAGCGACATCGTAAACCCGATTCATGCCACCTCG GTAATCGACAGGGAAATCGCTGATAAGCAGCTTCCGGCCCTAACCGCTGACGGTGCACCGCTCGGTAACGTCAGC CGGAAACCATGGCGATAACGGCCCCAGATAACCCGCTTGCAGTGCAGCTCC	9120 9200 9280

Figure 3

YiaJ-Ko	MGT -----KE	SEN TQDKERP	AGSQS LFRGL	M LIE ILSNYP	NGCPVAH LSE
YiaJ-Ec	MGKEVMGKKE	NEMAQEKERP	AGSQS LFRGL	M LIE ILSNYP	NGCPLAH LSE
YiaJ-Hi	MN IEVK ----	---MEKEKS	LGNQALIRGL	R ELD ILSNYP	NGCPLAK LAE
YiaJ-Ko	LAG LNKSTVB	RLLOGL OSCG	YVTPAPAAGS	VALTTKF IRV	GOKALSSLN I
YiaJ-Ec	LAG LNKSTVB	RLLOGL OSCG	YVTTPAPAAGS	YRLTTKF IAV	GOKALSSLN I
YiaJ-Hi	LAN LNKSTAB	RLLOGL QNEG	YVKPANAAGS	YRLTIKCLS I	GOKVLSSMNI
YiaJ-Ko	IHVAAPHL EA	LNLATGETVN	FSSREDDHAT	L TYKLEPTTG	MLRTRAY IGO
YiaJ-Ec	IHIAAPHL EA	LNIA TGETIN	FSSREDDHAT	L TYKLEPTTG	MLRTRAY IGO
YiaJ-Hi	THVASPY LEQ	LNLK LGETIN	FSKREDDHAT	M TYKLEPTNG	MLKTRAY IGO
YiaJ-Ko	HMR --CTARQ	WAKIYMAFGH	P-DYVESYWN	SHOE IIQPLT	RNTITGLPAM
YiaJ-Ec	HMP LYCSAM -	-GRKIYMAFGH	P-DYVKSYWE	SHOHE IQPLT	RNTITELPAM
YiaJ-Hi	YLK LYCSAM -	-GRKIFLAYEK	KVDYLSHYWQ	SHORE IKKLT	RYTITELDD I
YiaJ-Ko	HDELAQIRER	NMAMDREENE	LGVSCLA PV	FDIHGRVPYA	IS IS LS TSRL
YiaJ-Ec	FDELAHIRES	GAAMDREENE	LGVSCIAPV	FDIHGRVPYA	VS IS LS TSRL
YiaJ-Hi	KLELETIQT	AYAMDREENE	LGVTCIACP I	FDSFGQVEYA	ISVSMS IYRL
YiaJ-Ko	KQVGEKNLLK	PLRD TAE A IS	REL GF SVREG	-----	
YiaJ-Ec	KQVGEKNLLK	PLRE TAQA IS	NELGFTVRDD	LGA IT	
YiaJ-Hi	NKFG TDAFLQ	E IRK TAEQ IS	LELGYEN ---	----I	

• Figure 4 •

giaK-Ko	MKVTFEQLKE	A FN RVLLDAC	VARETADACA	EMFAR TTESG	VYSEGVNRP
giaK-Ec	MKVTFEQLKA	A FN RVLLISRG	VDSETADACA	EMFAR TTESG	VYSHGVNRP
giaK-Hi	MRVSYDELRN	E EKRVYLLDRQ	LTEELAEECA	TAFTD TTQAG	AYSEG INREP
giaK-Ko	R FIOOLDNGD	I IPEAOOPQRV	TTLGAIEQWD	AQRS IGN LTA	KKMMMDRAIEL
giaK-Ec	R FIQQLENGD	I IPDAQPKR I	TS LGAIEQWD	AQRS IGN LTA	KKMMMDRAIEL
giaK-Hi	R FIQQLEQGD	IVPNAIPTKV	LS LGS IEQWD	AHQAGIGNLTA	KKMMMDRAIEL
giaK-Ko	ASDHEGIGLVA	LRNANHWMRG	G SYGWQAAEK	GY IG ICWTNS	IAVMAPWGAK
giaK-Ec	AADHGIGLVA	LRNANHWMRG	G SYGWQAAEK	GY IG ICWTNS	IAVMPPWGAK
giaK-Hi	ASQHGVGV IA	LRNANHWMRG	G SYGWQAAEK	GY IG ICWTNA	LA VMP PWGAK
giaK-Ko	ECRIGTNPLI	VA IPS TP ITM	VDMMSMSMF SY	GM LE PVNR LAG	RE LPV DGG FD
giaK-Ec	ECRIGTNPLI	VA IPS TP ITM	VDMMSMSMF SY	GM LE PVNR LAG	RO LPV DGG FD
giaK-Hi	ECRIGTNPLI	IAVPTTP ITM	VDMSCSMYSY	GM LEVER LAG	RO TFV DAG PD
giaK-Ko	DDGR LTKEPG	TIEKNRR ILP	M GYWKGSGLS	IV LDM IATLL	SNGSSVAEV
giaK-Ec	DEGN LTKEPG	VIEKNRR ILP	M GYWKGSGMS	IV LDM IATLL	SDGASVAEV
giaK-Hi	DEGN LTRDPS	IVEKNRR LLP	M GFWKGSGLS	IV LDM IATLL	SNGESTVAV
giaK-Ko	QENSDEYGVS	Q IF IA IEVDK	L IDGATRDAK	L Q RIMDF ITT	AERADENVAV
giaK-Ec	QDNSDEYG IS	Q IF IA IEVDK	L IDGP TRDAK	L Q RIMDYV TS	AERADENQA I
giaK-Hi	EDKNDEYCVS	QVF IA IEVDR	L IDGKS KDEK	L NR IM DYVK T	AERSDP TOAV
giaK-Ko	R LPGHE FTRL	LDENRRNG IT	V DDSVWAK IQ	AL	
giaK-Ec	R LPGHE FTTL	LAENRRNG IT	V DDSVWAK IQ	AL	
giaK-Hi	R LPGHE FTT I	LSDNQ TNG IP	V DERYWAKLK	TL	

Figure 5

yiaL-Ko	M IFGB IAQPN	-PCRLPAA IE	RALDFLR TTD	FHALAPGVVE	IDGQN IFAQV
yiaL-Ec	M IPGH IAQPN	-PCRLPAA IE	RALDFLR ATD	FNALEPGVVE	IDGKN IY TQ I
yhch-Hi	M ISSLTNPN	FRVGLPKV IA	EVCDY LNTLD	LNALENGRHD	INDQ - IY MNV

yiaL-Ko	ID LTTRDAAE	NRPEVHRRYL	DIQFLASGEE	K IG IA ID TGN	NQ ISES LLEQ
yiaL-Ec	ID LTTRAEAVV	NRPEVHRRYI	DIQFLAWGEE	K IG IA ID TGN	NKVSES LLEQ
yhch-Hi	MEPE TAEPS	KKAELHBEYL	DVQVLIRGTE	N IEVGATYPN	LSKYEDYNEA

yiaL-Ko	RD IIFYHDSE	HESFFEMTPG	NYA IFFPQDV	HRPGCNKTV	TP - IRK IVVK
yiaL-Ec	RN IIFYHDSE	HESF IEM IPG	SYA IFFPQDV	HRPGC IMQTA	SE - IRK IVVK
yhch-Hi	DDYQLCADID	DKFTV TMKPK	MFAV FYPYEP	EKPCCVVNGK	TEK IKK LVVK

yiaL-Ko	VA ISVL-
yiaL-Ec	VALTA LN
yhch-Hi	VPVK - LI

yiaQ-Ko	MSRPLLQLAL	DHTSLQAAQR	DVALLDQDHVD	IVEAGTILCL	TEGLSAVKAL
yiaQ-Ec	MSRPLLQLAL	DHSSLEAAQR	DVTLLKDSVD	IVEAGTILCL	NEGLGAVKAL
yiaQ-Hi	MGKPLLQIAL	DAQYLETALV	DVKQIEHNID	TIEVGTILAC	SEGMRAVRIL

yiaQ-Ko	RAQCPGKIV	ADWKVADAGE	TLAQQAEGAG	ANWMTIICAA	P LATVEKGHA
yiaQ-Ec	REQCPDRIV	ADWKVADAGE	TLAQQAEGAG	ANWMTIICAA	P LATVEKGHA
yiaQ-Hi	RALYPNQILV	CDLKTTDAGA	TLAKMAFEAG	ADWLTVSAAA	HPATKAACQK

yiaQ-Ko	VAAQ-----	---CGGEIQM	E LFGNWTLDD	ARAWYRTGVH	QA IYBGRGRDA
yiaQ-Ec	MAQ-----	---CGGEIQI	E LFGNWTLDD	ARDWHRIGVR	QA IYBGRGRDA
yiaQ-Hi	VAAEFNKIQP	NLGVPKEIQI	E LYGNWNFDE	VKNWLQLGIK	QA IYBRSRDA

yiaQ-Ko	QASGQQWGEA	DLARMKALSD	IGLELSITGG	ITPADLPFLK	D IN -VKAFIA
yiaQ-Ec	QASGQQWGEA	DLARMKALSD	IGLELSITGG	ITPADLPFLK	D IR -VKAFIA
yiaQ-Hi	ELSGLWSNSQ	DIENIEKLDS	LGIELSITGG	ITPDLDLHLFK	NTKNLKAFIA

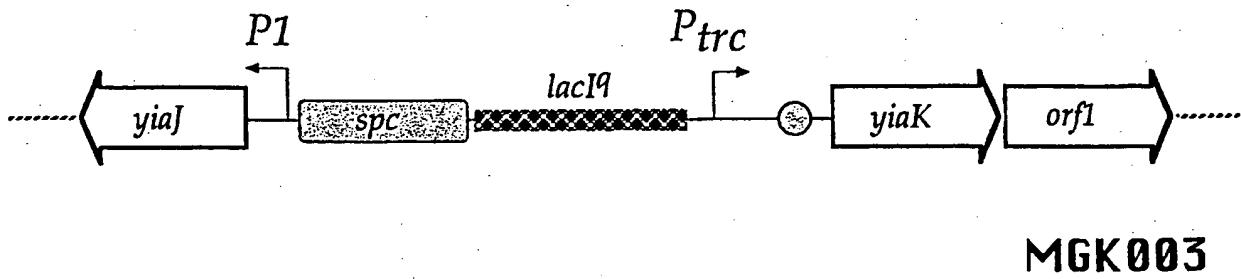
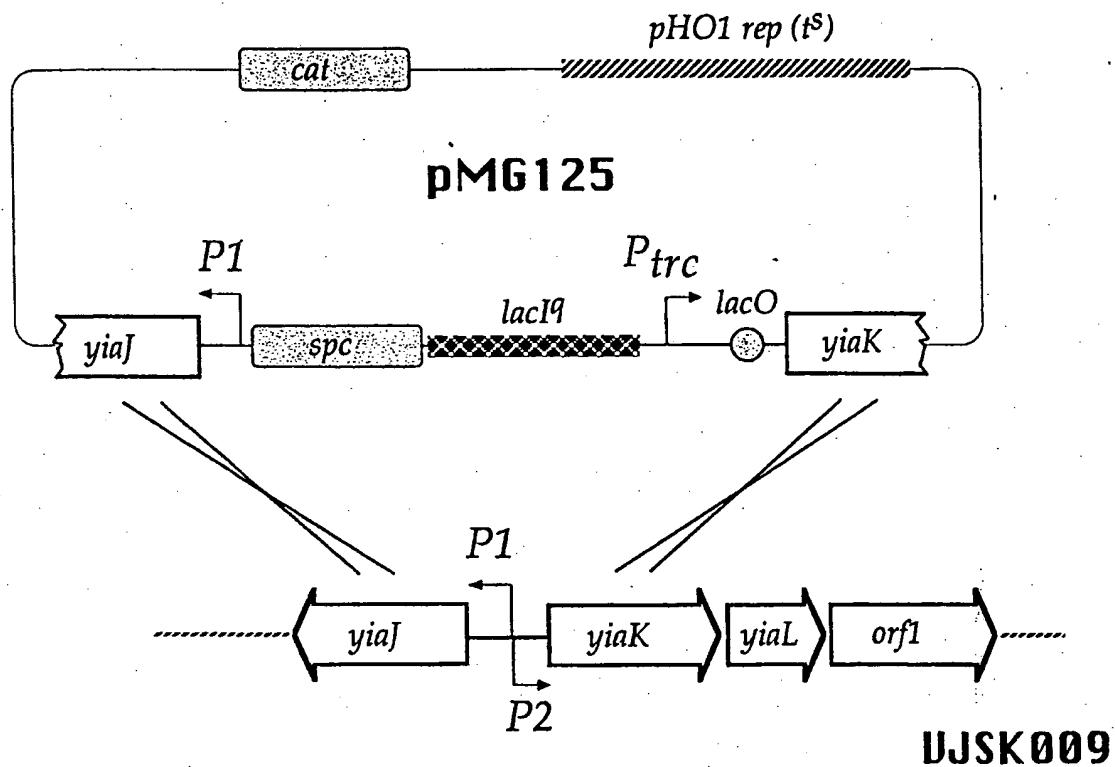
yiaQ-Ko	GRALAGAAHP	ARVAAEFHAQ	IDA IWGEQHA	
yiaQ-Ec	GRALAGAANP	AQVAGDFHAQ	IDA IWGGARA	
yiaQ-Hi	GRAVGKSGR	-EIAEQLKQK	IGQFW----I	

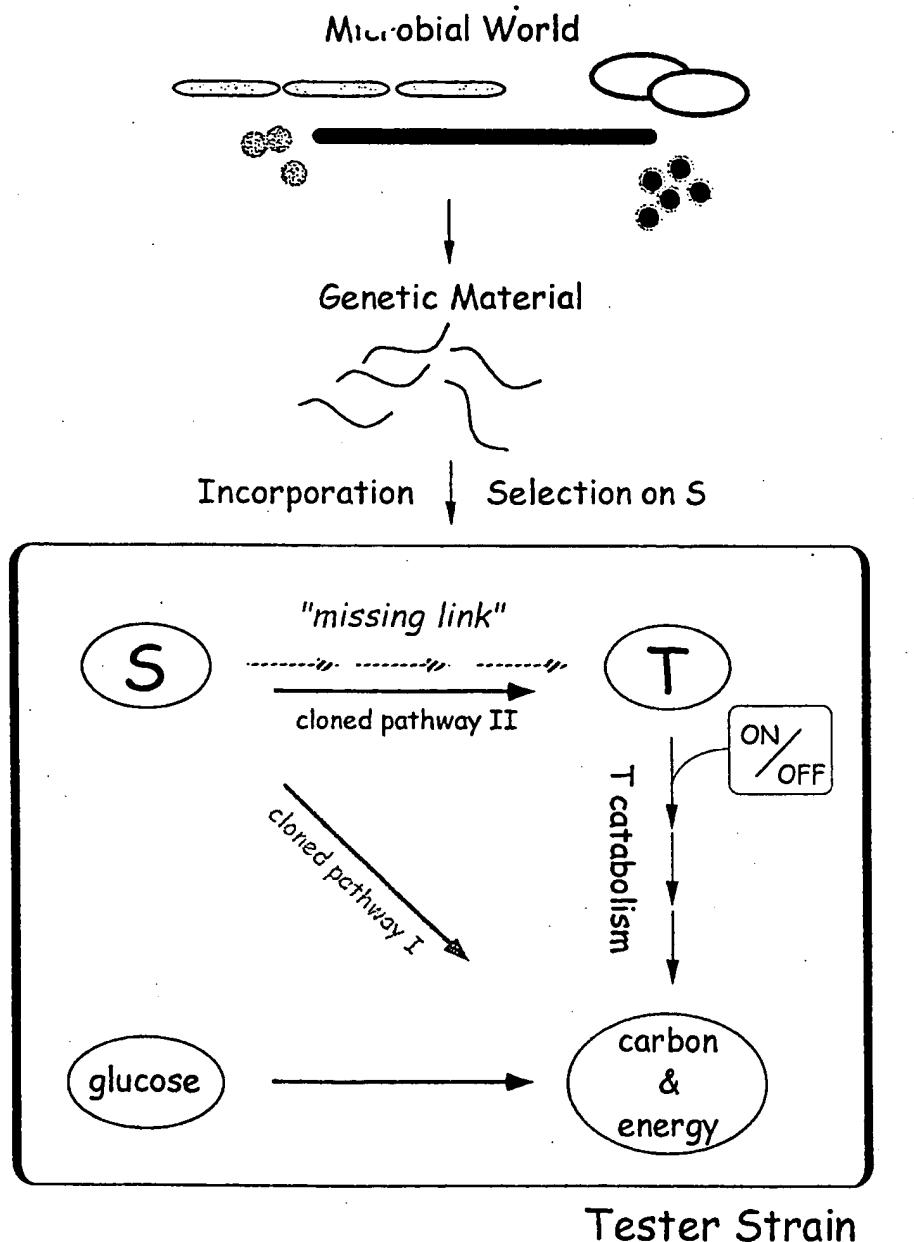
yiaR-Ko	MR -----	---NHP LG IV	EKA LAKD LSW	PER LV LAKSC	GFD FVEM SVD
yiaR-Ec	MRK STLSGEV	RVRNH QLG IV	EKA LAKD LSW	PER LV LAKSC	GFD FVEM SVD
yiaR-Hi	MKK -----	---EK IG IV	EKA LP KN ITW	QER LS LAKAC	GFE EIEMS ID
yiaR-Ko	ETDER LSRLE	WTPAQRAS LV	SAM LE TAVA I	PSMCLSAHRR	FFFGSRDEAV
yiaR-Ec	ETDER LSRLD	WSAAQRTS LV	AAM IE TGVG I	PSMCLSAHRR	FFFGSRDEAV
yiaR-Hi	ESNDRLSRLN	WTKSER IALH	QS I IQSG ITI	PSMCLSAHRR	FFFGSKDKK I
yiaR-Ko	RDRARE IMTR	AIR LARD LG I	RT IO LAGYDV	YYEEH DEGTR	QRFAEGLAWA
yiaR-Ec	RERARE IMSK	AIR LARD LG I	RT IO LAGYDV	YYEDH DEGTR	QRFAEGLAWA
yiaR-Hi	RQKS FE IMER	AID LSVN LG I	RT IO LAGYDV	YYEKQDEETI	KYFQEGIEFA
yiaR-Ko	VEQAAAQVM	LAVE IMD TAF	MNS ISKWKKW	DEM LSS PWFT	VYPDVG NLSA
yiaR-Ec	VEQAAAQSVM	LAVE IMD TAF	MNS ISKWKKW	DEM LAS PWFT	VYPDVG NLSA
yiaR-Hi	VTLLAASAQVT	LAVE IMD TPF	MSS ISRWKWW	DTI IN SPWFT	VYPDIGNLSA
yiaR-Ko	WGNDV TAE LK	LG IDR IAA IH	LKD TL PV TDD	SPGQ FR DVPF	GEGCVDFVG I
yiaR-Ec	WGNDV PAELLK	LG IDR IAA IH	LKD TO PV TGQ	SPGQ FR DVPF	GEGCVDFVG I
yiaR-Hi	WNNN IEEELT	LG>IDK IS A IH	LKD TYPV TET	SKGQ FR DVPF	GOGCVDFVHF
yiaR-Ko	FKTIRE LNYR	GSFL IEM WTE	KASEPV LE II	QARRW IESRM	QE GGFTC
yiaR-Ec	FKTLHK LNYR	GSFL IEM WTE	KAKEPV LE II	QARRW IEARM	QEAGF IC
yiaR-Hi	FS LLKK INYR	GAFL IEM WTE	KNEEPV LE II	QARKW IVQQM	EKAG LLC

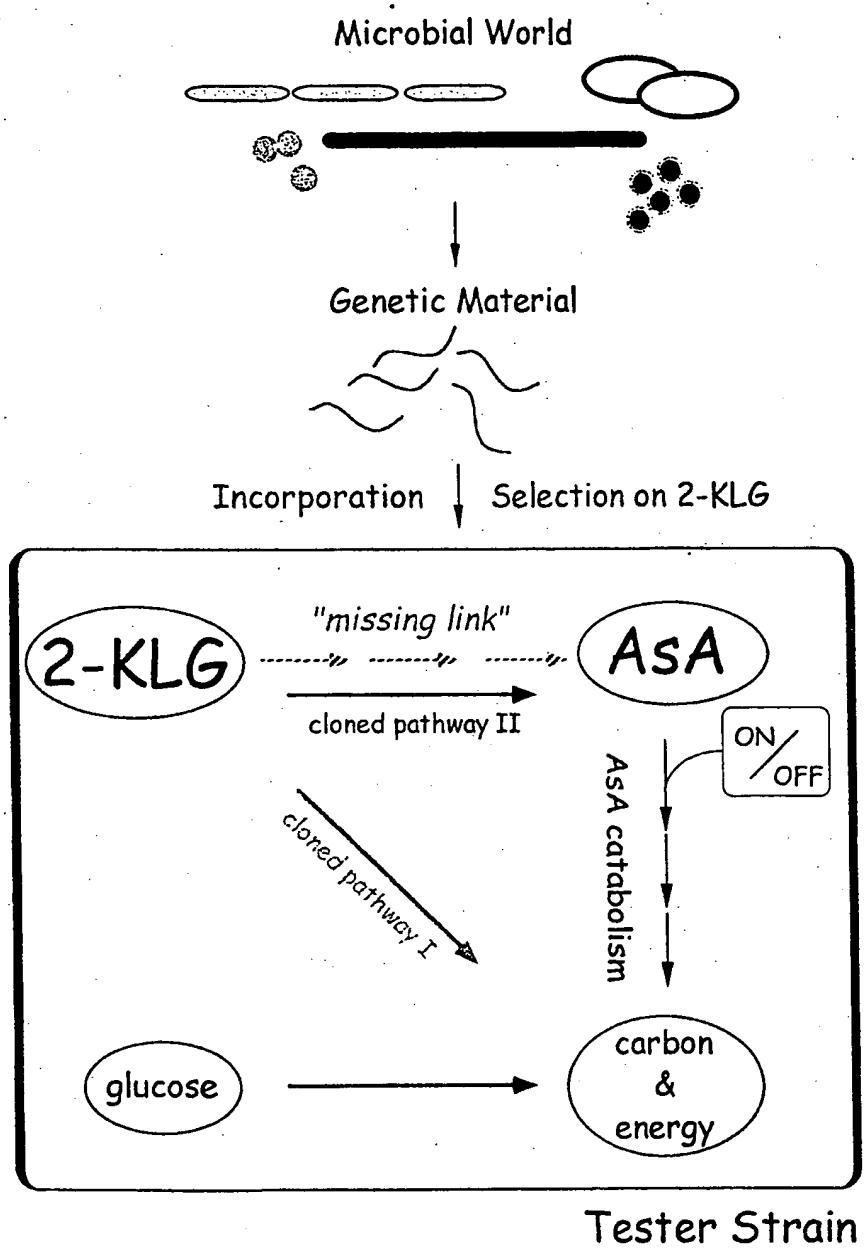
Figure 9

yiaS-Ko	MLEQLKAEVL	AANLA LPAAHG	LVTFTWGNSVS	AVDE TRKLMV	IKPSGV EYEV
yiaS-Ec	MLEQLKADVL	AANLA LPAAHH	LVTFTWGNSVS	AVDE TRQWMV	IKPSGV EYDV
yiaS-Hi	MLAQLKEVF	EANLA LPKEHH	LVTFTWGNSVS	ALDREKNLUVV	IKPSGV DYDV
yiaS-Ko	MTADDMVVVE	MASGVVVEGG	KKPSSDTPTTH	LA LYRRYPQI	GG IVH THSRH
yiaS-Ec	MTADDMVVVE	IASGVVVEGS	KKPSSDTPTTH	LA LYRRYAEI	GG IVH THSRH
yiaS-Hi	MTENDMVVVD	LFTGN IVEGN	KKPSSDTPTTH	LE LYRQFPHI	GG IVH THSRH
yiaS-Ko	ATIWSQAGLD	LPAWGTTHAD	YFYGA IPCTR	RMTVEE INGE	YEYQTGEVII
yiaS-Ec	ATIWSQAGLD	LPAWGTTHAD	YFYGA IPCTR	QMTAEE INGE	YEYQTGEVII
yiaS-Hi	ATIWAQAGLD	IIEVGTTGHD	YFYGTIPCTR	QMTTKE IKGN	YELETGKVIV
yiaS-Ko	KTFEQRGLDP	AQIPAVLVHS	HGPFAWGKDA	ADAVHNAVVL	EECAYMGLFS
yiaS-Ec	ETFEERGRSP	AQIPAVLVHS	HGPFAWGKNA	ADAVHNAVVL	EECAYMGLFS
yiaS-Hi	ETFLSRGIEP	DNIPAVLVHS	HGPFAWGKDA	NNAVHNAVVL	EEVAYMN LFS
yiaS-Ko	RQW -POLPDM	QSELLDKHYL	RKHGANAITG	KTSPAELPG	
yiaS-Ec	RQLAPOLPAM	QNELLDKHYL	RKHGANAYYG	Q -----	
yiaS-Hi	QQLNRYLSPM	QKDLLDKHYL	RKHGANAYYG	Q -----	

Figure 10

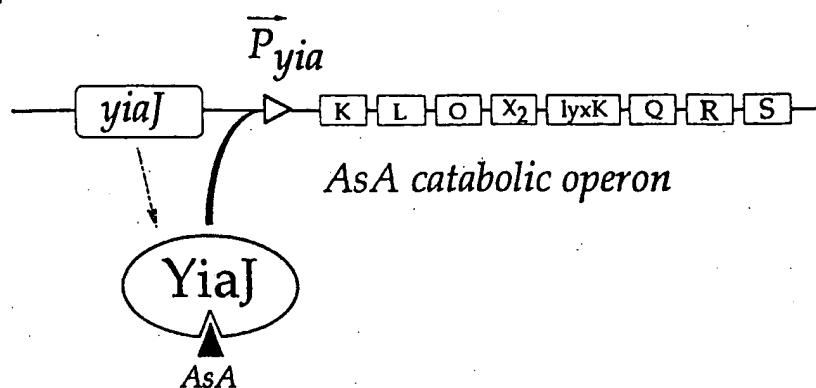
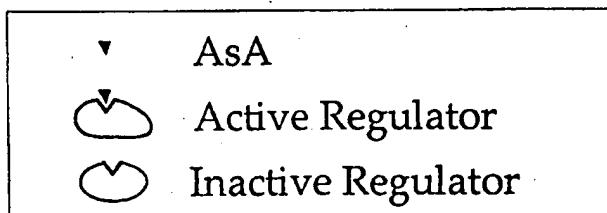
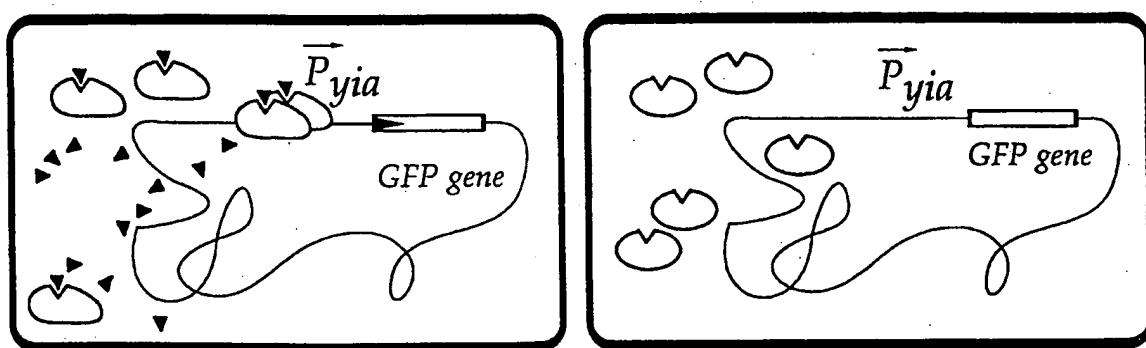






The Metabolic Selection Strategy

Figure 13

A**B**

SEQUENCE LISTING

<110> Hoch, James
Dartois, Veronique

<120> METABOLIC SELECTION METHODS

<130> WESLEY B. AMES: Microgenomics

<140>
<141>

<160> 33

<170> PatentIn Ver. 2.0

<210> 1
<211> 816
<212> DNA
<213> yia j

<400> 1
atgggcacaa aagaaagcga gaacacgcaa gataaagaga ggcctgccgg aagtcaagac 60
cttttcgtg gggttatgtct aatttagatc ctgagtaatt atccaaatgg ctgtcccg 120
gcatctgt cgaaactggc gggactgaac aaaagtaccg ttcatcgctt attacagggg 180
ctgcagtctt gcccgtacgt gacgcctgcc cccgcggcgg ggagctatgc gctgacgaca 240
aaatttatcc gcgttgccca aaaggcggtt tcgtcgctga atattatcca cgtcgccggc 300
ccgcattttt aggcgtttaa cctggccacc ggcgagacgg tgaacttctc cagccgtgaa 360
gatgaccacg cgatccgtat ttataagctg gagccgacca ccgttatgtct gctgacgac 420
gcctatattt gccagcacat gcgtgtact gctcgcaat gggcaagat ttatatggcg 480
tttggccatc ctgactacgt tgagagctac tggattcac accaggagat tatccagccg 540
ctgaccctgtt ataccattac cggcttgcct gcatgcattt atgaactggc gcatgatccgc 600
gagcggaaaata tggcgatgga cagggaagag aacgagctgg gcgtgtcg 660
cccggtttt atatccatgg ggcgtgcct tatgccattt ctatctctt atcaacatcg 720
cgccctcaagc aggtgggaga gaaaaattta ctcaagccgc tacgcatac ggcagaggcg 780
atttctcgcg aactggcattt ttcgtgcgg gaaggt 816

<210> 2
<211> 996
<212> DNA
<213> yia k

<400> 2
atgaaagtca cgttttagca gttaaaagag gcattcaatc gggtactgct ggacgcgtgc 60
gtcgccccggg aaaccgcga tgcctgcga gaaatgtttt cccgcaccac cgaatccggc 120
gtctattctc acggcgtgaa ccgtttccct cgcttcatcc agcagttgga taacggcgac 180
attatccctg aggctcaacc gcagcgggtt accacgctcg gcgcctatcg aacagtggat 240
gctcagcgtt ccatcgccaa cctgacggcg aaaaagatga tggatcgggc cattgagctg 300

gcctccgatc acggtatcg cgctggcgcc ttacgtaatg ctaaccactg gatgcgcggc 360
 ggcagctacg gctggcaggc ggccggaaaaa ggctacatcg gtatctgtg gaccaactcc 420
 atcggcgta tggcgccatg gggcgctaaa gagtgccgta tcggtaccaa cccgctgatc 480
 gtcgccattc cgtcgacgccc gatcaccatg gtggatatgt cgatgtcgat gtttcctac 540
 ggcatgctgg aggttaaccc ccttgcggc cgccgaaactgc ccgtggacgg cggattcgac 600
 gatgacggtc gtttgcacaa agagccgggg acgatcgaga aaaatcgccg cattttaccc 660
 atgggctact ggaaagggtc cggcctgtcg atcgtgctgg atatgattgc caccctctc 720
 tccaacggat cgtcggttgc cgaagtgacc caggaaaaca gcgatgaata tggcggttc 780
 cagatcttca tcgctattga agtgataag ctgatcgacg ggcaccccg cgacgccaag 840
 ctgcaacggaa ttatggatt catcaccacc gccgagcgcc ccgatgaaaaa tgtggcggtc 900
 cgtttctg gccatgaatt tacccgtctg ctggatgaaa accgcccaca cggcattacc 960
 gtcgatgaca gcgtatggc caaaatttagt ggcgtc 996

<210> 3
 <211> 462
 <212> DNA
 <213> yia 1

<400> 3
 atgatttttgcgtcatattgc tcaacctaattccgtgtcgctg tgcccgccgc cattgagcgg 60
 ggccttgatttccctgcgacac gacggatttc cacgcgtgg caccgcgt cgtggaaatc 120
 gacggccaaa acatcttcgc gcaggttatac gacttaacca ctcgcgtatgc cgctgaaaat 180
 cgtccggagg tccaccgtcg ctatctggat atccagtttc tggcatcgcc cgaagaaaaa 240
 atcggatcgccattgtatcg ccattgatac cggcaataat caaatcagcg aatctttatt agaacagcgc 300
 gatattatttttatcacga cagcgaacat gaatcggttct ttgaaatgac gccaggcaac 360
 tatgcgatattttcccgca agatgttcat cgtcctggat gtaataaaac tgtagccacg 420
 ccgatccgca aaatagtctgtaaaatcgct atttcagttt ta 462

<210> 4
 <211> 945
 <212> DNA
 <213> orfl

<400> 4
 atgaattcga ataataccgg ttacattatc ggtgcgtacc cctgtgcccc ctgtgcaccc 60
 tcatttcacc aaaagagtga agaggaagag atgaaattct ggcggcagct ctccgacacc 120
 ccggatattc gcgggctgaa gcaaccctgc ctaccctgccc ttgaacatct tcattccgtc 180
 ggcgacgagt ggttattgcg ccatacccg ggacactggc agattgtcgat taccgcaccc 240
 atggaaacca tgcgcgcgc cggtaaaac ggcggctttg ggctggcgtc cagcgacgaa 300
 acgcagcgca aagcctgcgt ggagtactat cgccacctgc agcagaagat cgctaaaatc 360
 aatggcaata ccgcccggaaa ggtcattgccc ttgagcttc acgcccggcc gctggcggtc 420
 aatgccaaacg tggctcaggc taccgacgcc ttgcccgtt cattaaaaga aattaccgc 480
 tggacttgtt cctgcgagct ggtgctggag cactgcgacg cgatgaccgg cagcgccgc 540
 cgcggaaaggat ttgtccgtt agaaaaacgtg ctgaaagcca ttgcccgtt tgacgttggc 600
 atttgttata actgggcgcg ttgcggcatt gaagggcgga ataccgtgct accgctcacc 660
 catacgcagc aggtaaaacg ggcaggaaag ctcggcgcgc tgatgtttc tggcacgacg 720
 cagaccggcg agtacggcga atggcaggat ttacacgcgc cgttgcgcgc ttctgcggc 780
 cagagcctga tgaccaccga acacgctcgta gattatttg cctgcgcagg aaccggcccc 840

ctgcaatttt caggcattaa attactggaa attaatgccs gcgcaaacgt tgatcatcgc 900
 atcgcgatat tacgcgacgg catctccgcg ctaaaaacaag cacaa 945

<210> 5
 <211> 1317
 <212> DNA
 <213> yia x2

<400> 5
 atgaatataa cctctaactc tacaaccaa gatataccgc gccagcgcgtg gttaagaatc 60
 attccgccta tactgatcac ttgttattt tcttatatgg accgggtcaa tattgccttt 120
 gcgatgcccgg gaggtatggc tgccgactta ggtatttccg ccaccatggc ggggctggcg 180
 ggcggatttt ctcttattcgg ttatctattt ttacaggttc cggcgggaa aattgccgtt 240
 cacggtagcgt gtaaagaatt tatcggttgg tcgctggcgt cctggggcgtt catctccgtg 300
 ctgacggggt taattaccaa tcagtaccag ctgctggccc tgcgcttctt actggggcgtg 360
 gcgaaaggcg gtagtgcgttgc ggtcggttctc acgatgatca gtaactgggtt cccccgacgt 420
 gaacgcggcgc ggcggcaacgc gattgtcatt atgtttgtgc cgattggcggg gattatcacc 480
 gccccactct caggcgtggat tatcacgggtt ctgcgactggc gctggctgtt tattatcgaa 540
 ggtttgcgttctc cgctgggttgc tctgggttgc tgggcataca ccattatgc cggccggcag 600
 gaagcgcgtt ggatttccga agcagagaag cgctatctgg tcgagacgtt gggccggag 660
 caaaaagcca ttgcggcgcac cgagggtaaaa aacgcctctc tgagcggcgt tctctccgac 720
 aaaaccatgt ggcagcttac ccgcctgaac ttcttctacc agaccggcat ttacggctac 780
 accctgtggc taccacccat tctgaaaagaa ttgaccacata gcagcatggg gcagggtggc 840
 atgcttgcca ttctgcccgtt cgtggcgcctt attgctgggat tggctgtt ttcctccctt 900
 tcagacccgaa ccggtaaaacg caagctgttc gtctgcctgc cgctgattgg ctgcgtctg 960
 tgcattttcc tgcgggtggc gctggggggaa caaatttggc ttcctatgc cgcgtggc 1020
 ggctgcggat tcttcctgc atcggcggctt ggcgtgttctt ggaccatccc ggcacgtctg 1080
 ttcagcgcgg aaatggcggg cggcgcgcgc ggggttatca acgcgttgg caacctcggc 1140
 ggatttgttgc gcccattatgc ggtcggttgc ctgatcacgt tgcgttgcgaa agacgtggc 1200
 gtctatttgcg tggcgatctc cctggcgcgtt gcccgcgttgc tggcgctgtt gctggccggcg 1260
 aaatgcgatg ccgggtgtgc gcccgttgcgaa acgataaaatc cacataaaacg cactgcg 1317

<210> 6
 <211> 1503
 <212> DNA
 <213> lyxk

<400> 6
 atgagcaaga aacaggcctt ctggctgggtt attgatttgcg gggcaccta tctgaaagcc 60
 ggtttatatg acgcccgaagg tcatgaacat ggcatttgcg ggcaagcgtt acggacgtt 120
 tcgccccctgc cgggttacgc cgaacgcgcac atgcgcgcac tctggcaaca ctgcgcggcg 180
 accatttgcg ggctatttaca gcaggcagggtt gtatccggcg aacagattaa aggctggggc 240
 atctccgcctc agggtcaagg gcttttctc ctcgataagc agatcggcc gctgggttac 300
 gccatccctc cctccgatcg tggcgctgtt ggggttatca acgcgttggca gggggaccgt 360
 attcccgaac ggctctatcc ctggctgggtt cagacgttgtt ggaccggaca tccggcttct 420
 ttgctgcgtt gggtaaaaga gaatgaaccc cagcgctacg cgcaattgg ctgcgtatg 480
 atggggcatg actatctgcg ctggctgttgc accggcgcgttgc agggctgcgtt gggagac 540
 atctccgatgtt ccaaccctcta caacatggcc atggggccgtt acgaccgcgtt cctgaccgcgtt 600

tggctggca tcggtaaat cgatagcgcg ctgccccccg ttgttagggtc agccgaaatt 660
 tgcggggaga tcaccgctca ggcagccgct ttaaccggtc tggcggcggt tactcccg 720
 gttggggccc tggttgcgt ggtctccacc gcccggcgcc ccgggattga ggatgagtcg 780
 accctcaatg cggtgatggg gacctgggccc gtcaactagcg gtatcgctca cggcctgc 840
 gaccatgagg cccaccctta cgtctatggc cgctacgtca atgacggcca gtatatcg 900
 cacgaagcca gcccggaccc atccggcaac ctgcgatggt ttaccggccca gtggggcgat 960
 ctctcggtt atgagatcaa tcaggccgctc gccagcctgc cgaaagccgg gagcgagctg 1020
 tttttctgc cggttctgtt tggcagcaac gcccggctgg agatgacctg cggctttac 1080
 ggcattgcagg cgctgcatac cccggccgac ctgctgcagg cggttatga aggctgttga 1140
 ttttagccata tgacccacct cagccgtatg cgcaacgc ttacaacgt tcaggccctg 1200
 cgcgtcaccg gcggcccgcc gcactccgac gtctggatgc agatgctggc ggacgtaagc 1260
 ggcttacgca ttgaactccc gaaggtggaa gagaccggct gttttggcgc ggcctcgcc 1320
 gctctgtcg gtaccggcgt ataccgcgc tttagcgaag cccggccgac cccggcagcac 1380
 cccggcgcga cgctgcgtcc cgatatgacc gcccacgcgc gctatcagcg caaataccgc 1440
 cactacctgc atttgcatttga agcaactacag ggctatcacg cccgtattaa ggagcacgc 1500
 tta 1503

<210> 7
<211> 660
<212> DNA
<213> yia q

<400> 7
atgagccgac cattactgca gctggcgctc gaccatacca gccttcaggc tgccgcagcgc 60
gatgtcgccc tgctacagga tcacgttgcatttgcgttggagg cggaaaccat cctctgc 120
accgaaggcc tttagccgtt taaagccctg cgcggccagt gtccggggaa gatcatcg 180
gccgacttggaa aagtgcggca cggccgtgaa accctggcgc agcaggccctt tggcgctggc 240
gccaacttggaa tgaccatcat ttgcggccca cccgcgtccca cggtcgagaa aggccacgccc 300
gtggcccccagg cctgcggccgg taaaatttgcgttggagctgt tcggcaactg gacgctggat 360
gacgcccgcg cctgggtaccc taccggcgctc catcaggcgatccatcg cggacgcgc 420
gcccaggccca gcggccgcgcgttggggggag gcggatctgg cgcgcattgaa agcgtgtcc 480
gatattggcc tttagctatc gattaccggc ggcatttgcgc cagccatctt acccgtgttc 540
aaagatatac acgtcaaagc ctttattgccc gggcgccgc tggcaggcgc ccccatccg 600
gcgcgggttgcgcgcaatttccacgcgtatc tctggggaga acagcatgc 660

<210> 8
<211> 858
<212> DNA
<213> yia r

<400> 8
atgcgttaacc acccggttagg tattttatgaa aaagcgctgg cgaaggatct cagctggcct 60
gagcggctgg tactggccaa aagctcggt ttgttgcgttgcgttggacgag 120
accgatgaac gccttcgcgc cctggagtgg acccccggccc agcgcgcacgc gctgggtgagc 180
gcgtatgtgg aaaccgcgtt cgcattccca tcgtatgtgcgttgtccca tcggccgttgc 240
cccttggca gcccgcgtatc agcggtacgc gatcggccgc gagagattat gaccaaaagcc 300
atccgcctgg cgcgcgtatc gggatccgc accatccgc tggcgggttgcgc gacgtctat 360
tacaaagacgatc atgatgaagg caccggcgatc cgttttgcggc aaggcgctggc ctggggcggttgc 420

gaacaggccg ccggccgcga ggtaatgctg gcgggtggaga tcatggacac cgccttatg 480
 aactccatca gcaaataggaa aaagtggac gagatgctt cgtcaccgtg gtttaccgtc 540
 tacccggacg tcggcaacct cagcgctgg ggaaacgacg tcaccggca gctgaagctg 600
 ggcatcgatc gtatcgccgc catccacctg aaagatacgc tgcccgtgac cgacgatagc 660
 cctggccagt tccgcacgt gccgttcggc gaaggatgcg tcgattttgt cggcatttt 720
 aagacgctgc gcgagctgaa ctaccgggt tcattttga ttgagatgtg gacggagaaa 780
 gccagcgagc cggtgctgga gattatccag gcccggcgat ggatcgaatc acggatgcag 840
 gaaggggat tcacatgt 858

<210> 9

<211> 714

<212> DNA

<213> yia s

<400> 9

atgttagaac aactgaaagc cgaggtactg gcggcaaacc tggccctccc cgcacacggc 60
 ctggtcacct ttacctgggg caacgtcagc gcggtcgatg aaacgcgcaa gctgtatggc 120
 attaagcctt cccgcgtcga atatgaggtg atgaccggc acgatatggt ggtcgtagag 180
 atggccagcg gtaaaagtcgt tgaaggcggt aaaaaaaccct cttcagatac gccaacgcatt 240
 ctggcgctt atcgcgccta tccgcagatc ggcgggatcg tgcataccca ctccccccac 300
 ggcacgatct ggtcgccaggc cgggctcgat ctcccccgcct ggggcaccac ccacggccac 360
 tacttctatg ggcgcgatccc ctgtacccga cggatgaccc ttgaggagat taacggcgag 420
 tatgagtatc agaccggcga ggtgattatc aaaacctttg aacagcgcgg cctggatccg 480
 ggcacaaatcc cggcggtatt ggtccattca cacggccctt ttgcctgggg taaagacgcc 540
 gcccacgccc tacataacgc cgtggtgctg gaggagtgcg cctacatggg cctcttcgc 600
 cggcagtggc cacagctgcc ggtatgcag tctgaactgc tcgataaaaca ctatctgcgt 660
 aaacacggcg cgaacgctat tacgggcaaa actagtcccg cggaaactcccc cggaa 714

<210> 10

<211> 272

<212> PRT

<213> YiaJ-Ko

<400> 10

Met	Gly	Thr	Lys	Glu	Ser	Glu	Asn	Thr	Gln	Asp	Lys	Glu	Arg	Pro	Ala
1															

5

10

15

Gly	Ser	Gln	Ser	Leu	Phe	Arg	Gly	Leu	Met	Leu	Ile	Glu	Ile	Leu	Ser
20															

25

30

Asn	Tyr	Pro	Asn	Gly	Cys	Pro	Val	Ala	His	Leu	Ser	Glu	Leu	Ala	Gly
35															

40

45

Leu	Asn	Lys	Ser	Thr	Val	His	Arg	Leu	Leu	Gln	Gly	Leu	Gln	Ser	Cys
50															

55

60

Gly	Tyr	Val	Thr	Pro	Ala	Pro	Ala	Ala	Gly	Ser	Tyr	Ala	Leu	Thr	Thr
65															

70

75

80

<210> 11
<211> 332
<212> PRT
<213> YiaK-Ko

<400> 11
Met Lys Val Thr Phe Glu Gln Leu Lys Glu Ala Phe Asp Arg Val Leu

1	5	10	15
Leu Asp Ala Cys Val Ala Arg Glu Thr Ala Asp Ala Cys Ala Glu Met			
20	25	30	
Phe Ala Arg Thr Thr Glu Ser Gly Val Tyr Ser His Gly Val Asn Arg			
35	40	45	
Phe Pro Arg Phe Ile Gln Gln Leu Asp Asn Gly Asp Ile Ile Pro Glu			
50	55	60	
Ala Gln Pro Gln Arg Val Thr Thr Leu Gly Ala Ile Glu Gln Trp Asp			
65	70	75	80
Ala Gln Arg Ser Ile Gly Asn Leu Thr Ala Lys Lys Met Met Asp Arg			
85	90	95	
Ala Ile Glu Leu Ala Ser Asp His Gly Ile Gly Leu Val Ala Leu Arg			
100	105	110	
Asn Ala Asn His Trp Met Arg Gly Gly Ser Tyr Gly Trp Gln Ala Ala			
115	120	125	
Glu Lys Gly Tyr Ile Gly Ile Cys Trp Thr Asn Ser Ile Ala Val Met			
130	135	140	
Ala Pro Trp Gly Ala Lys Glu Cys Arg Ile Gly Thr Asn Pro Leu Ile			
145	150	155	160
Val Ala Ile Pro Ser Thr Pro Ile Thr Met Val Asp Met Ser Met Ser			
165	170	175	
Met Phe Ser Tyr Gly Met Leu Glu Val Asn Arg Leu Ala Gly Arg Glu			
180	185	190	
Leu Pro Val Asp Gly Gly Phe Asp Asp Asp Gly Arg Leu Thr Lys Glu			
195	200	205	
Pro Gly Thr Ile Glu Lys Asn Arg Arg Ile Leu Pro Met Gly Tyr Trp			
210	215	220	
Lys Gly Ser Gly Leu Ser Ile Val Leu Asp Met Ile Ala Thr Leu Leu			
225	230	235	240
Ser Asn Gly Ser Ser Val Ala Glu Val Thr Gln Glu Asn Ser Asp Glu			
245	250	255	
Tyr Gly Val Ser Gln Ile Phe Ile Ala Ile Glu Val Asp Lys Leu Ile			

260

265

270

Asp Gly Ala Thr Arg Asp Ala Lys Leu Gln Arg Ile Met Asp Phe Ile
 275 280 285

Thr Thr Ala Glu Arg Ala Asp Glu Asn Val Ala Val Arg Leu Pro Gly
 290 295 300

His Glu Phe Thr Arg Leu Leu Asp Glu Asn Arg Arg Asn Gly Ile Thr
 305 310 315 320

Val Asp Asp Ser Val Trp Ala Lys Ile Gln Ala Leu
 325 330

<210> 12

<211> 154

<212> PRT

<213> YiaL-Ko

<400> 12

Met Ile Phe Gly His Ile Ala Gln Pro Asn Pro Cys Arg Leu Pro Ala
 1 5 10 15

Ala Ile Glu Arg Ala Leu Asp Phe Leu Arg Thr Thr Asp Phe His Ala
 20 25 30

Leu Ala Pro Gly Val Val Glu Ile Asp Gly Gln Asn Ile Phe Ala Gln
 35 40 45

Val Ile Asp Leu Thr Thr Arg Asp Ala Ala Glu Asn Arg Pro Glu Val
 50 55 60

His Arg Arg Tyr Leu Asp Ile Gln Phe Leu Ala Ser Gly Glu Glu Lys
 65 70 75 80

Ile Gly Ile Ala Ile Asp Thr Gly Asn Asn Gln Ile Ser Glu Ser Leu
 85 90 95

Leu Glu Gln Arg Asp Ile Ile Phe Tyr His Asp Ser Glu His Glu Ser
 100 105 110

Phe Phe Glu Met Thr Pro Gly Asn Tyr Ala Ile Phe Phe Pro Gln Asp
 115 120 125

Val His Arg Pro Gly Cys Asn Lys Thr Val Ala Thr Pro Ile Arg Lys
 130 135 140

Ile Val Val Lys Val Ala Ile Ser Val Leu
 145 150

<210> 13
 <211> 315
 <212> PRT
 <213> ORF1

<400> 13
 Met Asn Ser Asn Asn Thr Gly Tyr Ile Ile Gly Ala Tyr Pro Cys Ala
 1 5 10 15

Pro Cys Ala Pro Ser Phe His Gln Lys Ser Glu Glu Glu Met Glu
 20 25 30

Phe Trp Arg Gln Leu Ser Asp Thr Pro Asp Ile Arg Gly Leu Glu Gln
 35 40 45

Pro Cys Leu Pro Cys Leu Glu His Leu His Pro Leu Gly Asp Glu Trp
 50 55 60

Leu Leu Arg His Thr Pro Gly His Trp Gln Ile Val Val Thr Ala Ile
 65 70 75 80

Met Glu Thr Met Arg Arg Arg Gly Glu Asn Gly Gly Phe Gly Leu Ala
 85 90 95

Ser Ser Asp Glu Thr Gln Arg Lys Ala Cys Val Glu Tyr Tyr Arg His
 100 105 110

Leu Gln Gln Lys Ile Ala Lys Ile Asn Gly Asn Thr Ala Gly Lys Val
 115 120 125

Ile Ala Leu Glu Leu His Ala Ala Pro Leu Ala Gly Asn Ala Asn Val
 130 135 140

Ala Gln Ala Thr Asp Ala Phe Ala Arg Ser Leu Lys Glu Ile Thr Arg
 145 150 155 160

Trp Asp Trp Ser Cys Glu Leu Val Leu Glu His Cys Asp Ala Met Thr
 165 170 175

Gly Ser Ala Pro Arg Lys Gly Phe Leu Pro Leu Glu Asn Val Leu Glu
 180 185 190

Ala Ile Ala Asp Tyr Asp Val Gly Ile Cys Ile Asn Trp Ala Arg Ser
 195 200 205

Ala Ile Glu Gly Arg Asn Thr Val Leu Pro Leu Thr His Thr Gln Gln
210 215 220

Val Lys Arg Ala Gly Lys Leu Gly Ala Leu Met Phe Ser Gly Thr Thr
225 230 235 240

Gln Thr Gly Glu Tyr Gly Glu Trp Gln Asp Leu His Ala Pro Phe Ala
245 250 255

Pro Phe Cys Pro Gln Ser Leu Met Thr Thr Glu His Ala Arg Glu Leu
260 265 270

Phe Ala Cys Ala Gly Thr Ala Pro Leu Gln Phe Ser Gly Ile Lys Leu
275 280 285

Leu Glu Ile Asn Ala Ser Ala Asn Val Asp His Arg Ile Ala Ile Leu
290 295 300

Arg Asp Gly Ile Ser Ala Leu Lys Gln Ala Gln
305 310 315

<210> 14

<211> 439

<212> PRT

<213> YiaX2

<400> 14

Met Asn Ile Thr Ser Asn Ser Thr Thr Lys Asp Ile Pro Arg Gln Arg
1 5 10 15

Trp Leu Arg Ile Ile Pro Pro Ile Leu Ile Thr Cys Ile Ile Ser Tyr
20 25 30

Met Asp Arg Val Asn Ile Ala Phe Ala Met Pro Gly Gly Met Asp Ala
35 40 45

Asp Leu Gly Ile Ser Ala Thr Met Ala Gly Leu Ala Gly Gly Ile Phe
50 55 60

Phe Ile Gly Tyr Leu Phe Leu Gln Val Pro Gly Gly Lys Ile Ala Val
65 70 75 80

His Gly Ser Gly Lys Lys Phe Ile Gly Trp Ser Leu Val Ala Trp Ala
85 90 95

Val Ile Ser Val Leu Thr Gly Leu Ile Thr Asn Gln Tyr Gln Leu Leu

100	105	110
Ala Leu Arg Phe Leu Leu Gly Val Ala Glu Gly Gly Met Leu Pro Val		
115	120	125
Val Leu Thr Met Ile Ser Asn Trp Phe Pro Asp Ala Glu Arg Gly Arg		
130	135	140
Ala Asn Ala Ile Val Ile Met Phe Val Pro Ile Ala Gly Ile Ile Thr		
145	150	155
Ala Pro Leu Ser Gly Trp Ile Ile Thr Val Leu Asp Trp Arg Trp Leu		
165	170	175
Phe Ile Ile Glu Gly Leu Leu Ser Leu Val Val Leu Val Trp Ala		
180	185	190
Tyr Thr Ile Tyr Asp Arg Pro Gln Glu Ala Arg Trp Ile Ser Glu Ala		
195	200	205
Glu Lys Arg Tyr Leu Val Glu Thr Leu Ala Ala Glu Gln Lys Ala Ile		
210	215	220
Ala Gly Thr Glu Val Lys Asn Ala Ser Leu Ser Ala Val Leu Ser Asp		
225	230	235
Lys Thr Met Trp Gln Leu Ile Ala Leu Asn Phe Phe Tyr Gln Thr Gly		
245	250	255
Ile Tyr Gly Tyr Thr Leu Trp Leu Pro Thr Ile Leu Lys Glu Leu Thr		
260	265	270
His Ser Ser Met Gly Gln Val Gly Met Leu Ala Ile Leu Pro Tyr Val		
275	280	285
Gly Ala Ile Ala Gly Met Phe Leu Phe Ser Ser Leu Ser Asp Arg Thr		
290	295	300
Gly Lys Arg Lys Leu Phe Val Cys Leu Pro Leu Ile Gly Phe Ala Leu		
305	310	315
Cys Met Phe Leu Ser Val Ala Leu Lys Asn Gln Ile Trp Leu Ser Tyr		
325	330	335
Ala Ala Leu Val Gly Cys Gly Phe Phe Leu Gln Ser Ala Ala Gly Val		
340	345	350
Phe Trp Thr Ile Pro Ala Arg Leu Phe Ser Ala Glu Met Ala Gly Gly		

355

360

365

Ala Arg Gly Val Ile Asn Ala Leu Gly Asn Leu Gly Gly Phe Cys Gly
 370 375 380

Pro Tyr Ala Val Gly Val Leu Ile Thr Leu Tyr Ser Lys Asp Ala Gly
 385 390 395 400

Val Tyr Cys Leu Ala Ile Ser Leu Ala Leu Ala Ala Leu Met Ala Leu
 405 410 415

Leu Leu Pro Ala Lys Cys Asp Ala Gly Ala Ala Pro Val Lys Thr Ile
 420 425 430

Asn Pro His Lys Arg Thr Ala
 435

<210> 15

<211> 501

<212> PRT

<213> LyxK-Ko

<400> 15

Met Ser Lys Lys Gln Ala Phe Trp Leu Gly Ile Asp Cys Gly Thr
 1 5 10 15

Tyr Leu Lys Ala Gly Leu Tyr Asp Ala Glu Gly His Glu His Gly Ile
 20 25 30

Val Arg Gln Ala Leu Arg Thr Met Ser Pro Leu Pro Gly Tyr Ala Glu
 35 40 45

Arg Asp Met Arg Gln Leu Trp Gln His Cys Ala Ala Thr Ile Ala Gly
 50 55 60

Leu Leu Gln Gln Ala Gly Val Ser Gly Glu Gln Ile Lys Gly Val Gly
 65 70 75 80

Ile Ser Ala Gln Gly Gln Gly Leu Phe Leu Leu Asp Lys Gln Asp Arg
 85 90 95

Pro Leu Gly Asn Ala Ile Leu Ser Ser Asp Arg Arg Ala Leu Lys Ile
 100 105 110

Val Gln Arg Trp Gln Arg Asp Arg Ile Pro Glu Arg Leu Tyr Pro Val
 115 120 125

Thr Arg Gln Thr Leu Trp Thr Gly His Pro Ala Ser Leu Leu Arg Trp
 130 135 140

Val Lys Glu Asn Glu Pro Gln Arg Tyr Ala Gln Ile Gly Cys Val Met
 145 150 155 160

Met Gly His Asp Tyr Leu Arg Trp Cys Leu Thr Gly Ala Lys Gly Cys
 165 170 175

Glu Glu Ser Asn Ile Ser Glu Ser Asn Leu Tyr Asn Met Ala Met Gly
 180 185 190

Gln Tyr Asp Pro Arg Leu Thr Glu Trp Leu Gly Ile Gly Glu Ile Asp
 195 200 205

Ser Ala Leu Pro Pro Val Val Gly Ser Ala Glu Ile Cys Gly Glu Ile
 210 215 220

Thr Ala Gln Ala Ala Ala Leu Thr Gly Leu Ala Ala Gly Thr Pro Val
 225 230 235 240

Val Gly Gly Leu Phe Asp Val Val Ser Thr Ala Leu Cys Ala Gly Ile
 245 250 255

Glu Asp Glu Ser Thr Leu Asn Ala Val Met Gly Thr Trp Ala Val Thr
 260 265 270

Ser Gly Ile Ala His Gly Leu Arg Asp His Glu Ala His Pro Tyr Val
 275 280 285

Tyr Gly Arg Tyr Val Asn Asp Gly Gln Tyr Ile Val His Glu Ala Ser
 290 295 300

Pro Thr Ser Ser Gly Asn Leu Glu Trp Phe Thr Ala Gln Trp Gly Asp
 305 310 315 320

Leu Ser Phe Asp Glu Ile Asn Gln Ala Val Ala Ser Leu Pro Lys Ala
 325 330 335

Gly Ser Glu Leu Phe Phe Leu Pro Phe Leu Tyr Gly Ser Asn Ala Gly
 340 345 350

Leu Glu Met Thr Cys Gly Phe Tyr Gly Met Gln Ala Leu His Thr Arg
 355 360 365

Ala His Leu Leu Gln Ala Val Tyr Glu Gly Val Val Phe Ser His Met
 370 375 380

Thr His Leu Ser Arg Met Arg Glu Arg Phe Thr Asn Val Gln Ala Leu
 385 390 395 400

Arg Val Thr Gly Gly Pro Ala His Ser Asp Val Trp Met Gln Met Leu
 405 410 415

Ala Asp Val Ser Gly Leu Arg Ile Glu Leu Pro Lys Val Glu Glu Thr
 420 425 430

Gly Cys Phe Gly Ala Ala Leu Ala Ala Arg Val Gly Thr Gly Val Tyr
 435 440 445

Arg Ser Phe Ser Glu Ala Arg Arg Ala Arg Gln His Pro Val Arg Thr
 450 455 460

Leu Leu Pro Asp Met Thr Ala His Ala Arg Tyr Gln Arg Lys Tyr Arg
 465 470 475 480

His Tyr Leu His Leu Ile Glu Ala Leu Gln Gly Tyr His Ala Arg Ile
 485 490 495

Lys Glu His Ala Leu
 500

<210> 16

<211> 220

<212> PRT

<213> YiaQ-Ko

<400> 16

Met Ser Arg Pro Leu Leu Gln Leu Ala Leu Asp His Thr Ser Leu Gln
 1 5 10 15

Ala Ala Gln Arg Asp Val Ala Leu Leu Gln Asp His Val Asp Ile Val
 20 25 30

Glu Ala Gly Thr Ile Leu Cys Leu Thr Glu Gly Leu Ser Ala Val Lys
 35 40 45

Ala Leu Arg Ala Gln Cys Pro Gly Lys Ile Ile Val Ala Asp Trp Lys
 50 55 60

Val Ala Asp Ala Gly Glu Thr Leu Ala Gln Gln Ala Phe Gly Ala Gly
 65 70 75 80

Ala Asn Trp Met Thr Ile Ile Cys Ala Ala Pro Leu Ala Thr Val Glu
 85 90 95

Lys Gly His Ala Val Ala Gln Ala Cys Gly Gly Glu Ile Gln Met Glu
 100 105 110

Leu Phe Gly Asn Trp Thr Leu Asp Asp Ala Arg Ala Trp Tyr Arg Thr
 115 120 125

Gly Val His Gln Ala Ile Tyr His Arg Gly Arg Asp Ala Gln Ala Ser
 130 135 140

Gly Gln Gln Trp Gly Glu Ala Asp Leu Ala Arg Met Lys Ala Leu Ser
 145 150 155 160

Asp Ile Gly Leu Glu Leu Ser Ile Thr Gly Gly Ile Thr Pro Ala Asp
 165 170 175

Leu Pro Leu Phe Lys Asp Ile Asn Val Lys Ala Phe Ile Ala Gly Arg
 180 185 190

Ala Leu Ala Gly Ala Ala His Pro Ala Arg Val Ala Ala Glu Phe His
 195 200 205

Ala Gln Ile Asp Ala Ile Trp Gly Glu Gln His Ala
 210 215 220

<210> 17

<211> 286

<212> PRT

<213> YiaR-Ko

<400> 17

Met Arg Asn His Pro Leu Gly Ile Tyr Glu Lys Ala Leu Ala Lys Asp
 1 5 10 15

Leu Ser Trp Pro Glu Arg Leu Val Leu Ala Lys Ser Cys Gly Phe Asp
 20 25 30

Phe Val Glu Met Ser Val Asp Glu Thr Asp Glu Arg Leu Ser Arg Leu
 35 40 45

Glu Trp Thr Pro Ala Gln Arg Ala Ser Leu Val Ser Ala Met Leu Glu
 50 55 60

Thr Ala Val Ala Ile Pro Ser Met Cys Leu Ser Ala His Arg Arg Phe
 65 70 75 80

Pro Phe Gly Ser Arg Asp Glu Ala Val Arg Asp Arg Ala Arg Glu Ile

85

90

95

Met Thr Lys Ala Ile Arg Leu Ala Arg Asp Leu Gly Ile Arg Thr Ile
 100 105 110

Gln Leu Ala Gly Tyr Asp Val Tyr Tyr Glu Glu His Asp Glu Gly Thr
 115 120 125

Arg Gln Arg Phe Ala Glu Gly Leu Ala Trp Ala Val Glu Gln Ala Ala
 130 135 140

Ala Ala Gln Val Met Leu Ala Val Glu Ile Met Asp Thr Ala Phe Met
 145 150 155 160

Asn Ser Ile Ser Lys Trp Lys Lys Trp Asp Glu Met Leu Ser Ser Pro
 165 170 175

Trp Phe Thr Val Tyr Pro Asp Val Gly Asn Leu Ser Ala Trp Gly Asn
 180 185 190

Asp Val Thr Ala Glu Leu Lys Leu Gly Ile Asp Arg Ile Ala Ala Ile
 195 200 205

His Leu Lys Asp Thr Leu Pro Val Thr Asp Asp Ser Pro Gly Gln Phe
 210 215 220

Arg Asp Val Pro Phe Gly Glu Gly Cys Val Asp Phe Val Gly Ile Phe
 225 230 235 240

Lys Thr Leu Arg Glu Leu Asn Tyr Arg Gly Ser Phe Leu Ile Glu Met
 245 250 255

Trp Thr Glu Lys Ala Ser Glu Pro Val Leu Glu Ile Ile Gln Ala Arg
 260 265 270

Arg Trp Ile Glu Ser Arg Met Gln Glu Gly Gly Phe Thr Cys
 275 280 285

<210> 18

<211> 238

<212> PRT

<213> YiaS-Ko

<400> 18

Met Leu Glu Gln Leu Lys Ala Glu Val Leu Ala Ala Asn Leu Ala Leu
 1 5 10 15

Pro Ala His Gly Leu Val Thr Phe Thr Trp Gly Asn Val Ser Ala Val
 20 25 30

Asp Glu Thr Arg Lys Leu Met Val Ile Lys Pro Ser Gly Val Glu Tyr
 35 40 45

Glu Val Met Thr Ala Asp Asp Met Val Val Val Glu Met Ala Ser Gly
 50 55 60

Lys Val Val Glu Gly Gly Lys Pro Ser Ser Asp Thr Pro Thr His
 65 70 75 80

Leu Ala Leu Tyr Arg Arg Tyr Pro Gln Ile Gly Gly Ile Val His Thr
 85 90 95

His Ser Arg His Ala Thr Ile Trp Ser Gln Ala Gly Leu Asp Leu Pro
 100 105 110

Ala Trp Gly Thr Thr His Ala Asp Tyr Phe Tyr Gly Ala Ile Pro Cys
 115 120 125

Thr Arg Arg Met Thr Val Glu Glu Ile Asn Gly Glu Tyr Glu Tyr Gln
 130 135 140

Thr Gly Glu Val Ile Ile Lys Thr Phe Glu Gln Arg Gly Leu Asp Pro
 145 150 155 160

Ala Gln Ile Pro Ala Val Leu Val His Ser His Gly Pro Phe Ala Trp
 165 170 175

Gly Lys Asp Ala Ala Asp Ala Val His Asn Ala Val Val Leu Glu Glu
 180 185 190

Cys Ala Tyr Met Gly Leu Phe Ser Arg Gln Trp Pro Gln Leu Pro Asp
 195 200 205

Met Gln Ser Glu Leu Leu Asp Lys His Tyr Leu Arg Lys His Gly Ala
 210 215 220

Asn Ala Ile Thr Gly Lys Thr Ser Pro Ala Glu Leu Pro Gly
 225 230 235

<210> 19

<211> 9334

<212> DNA

<213> yia

<400> 19

ggatccgcgg ggcacaaaggc ggagacgcca gaacagtctt ggtcctgtc atggacacc 60
acgcaggcga cttcacagt acggcagccg atgcacttct ccgcattccgc gagaataaac 120
cgattcatcc ttctccattt gggataaaaaa cgcaagatgc cagaaaaaac ccgtttctt 180
ctcccttgc tcctgaatgg agtcagcggc gtttctctc agatgtccgg gattatctgg 240
tcatttgcct taacttccc gcacgaaaaa gcccagttc cgagaaatcg ccttgccgt 300
atcgcgtac ggcttgagta aattttctc tcccacctgc ttgaggcgcg atgtttagat 360
agagatagaa atggcataag gcacgcgccc atggatatac aaaacgggaa cagccaggca 420
cgacacgccc agtcgttct ctccccgtc catgccata ttcgctcgc ggatctgcgc 480
cagttcatca tgcatcgca gcaagccggta aatggtatta cgggtcagcg gctggataat 540
ctcctggtgt gaattccagt agctctcaac gtatcagga tggccaaacg ccatataaaat 600
cttgcggcat tgccgagcag tacagcgcata gtgtcgccat atataggcgc gcgtacgcag 660
cataccggtg gtcggctcca gcttataaat caggatcgcc tggtcatctt cacggctgga 720
gaagttcacc gtctcgccgg tggccaggtt aagcgcctca agatgcggcg ccgcgacgtg 780
gataatattt acgcacgaca acgcctttt gccaacgcgg ataaattttt tcgtcagcgc 840
atagctcccc gccgcgggg caggcgtcac gtacccgcag gactgcagcc cctgtaataa 900
gcatgtaacg gtactttgt tcaagccgc cagttccgac agatgcggca cgggacagcc 960
atttggataa ttactcagga tctcaatttag catcaaccca cggaaaaggc tctgacttcc 1020
ggcaggcetc tctttatctt gcgtttctc gctttctttt gtgcccacatcg cttccgcgtcc 1080
cattttgtc gcgttcagat ggtacgcac agtgtgttcc agttcacgt ctgaaccgaa 1140
aaaacacaac tttatgattt ttatgattt taaaataac gctgcccgtt gatctgacaa 1200
aaattgtatcg ctatatttga aatcagattt cgcatagtga aatttagaga taaaaaaggcg 1260
atcaactctg accagaaaaa cagcaatgaa agtcacgtt gaggcattaa aagaggcatt 1320
caatcgggta ctgctggacg cgtgcgtcgc ccggaaacc gccgatgcct ggcgcagaaat 1380
gtttccccgc accaccgaat ccggcgtcta ttctcacggc gtgaaccgct ttcctcgctt 1440
catccagcag ttggataacg ggcacattat ccctgaggct caaccgcagc ggggaccac 1500
gctcggcgcc atcgaacagt gggatgcctca ggcgtccatc ggcaacctga cggcgaaaaa 1560
gatgatggat cggccattt agtcggcetc cgatcacggt atcggcctgg tcgccttacg 1620
taatgtcaac cactggatgc gcccggcag ctacggctgg caggcggcgg aaaaaggcta 1680
catcggtatc tgctggacca actccatcgc cgttatggcg ccatggggcg ctaaagagtg 1740
ccgtatcggt accaaccgcg tgatgcgtc cattccgtcg acgcccgtca ccatgggtga 1800
tatgtcgatc tgatgttct cctacggcat gctggagggtt aaccgcctt ccggccgcga 1860
actgcccgtg gacggcgat tgcacgtga cggtcgtttt accaaagagc cggggacgat 1920
cgagaaaaat cggccattt taccatggg ctactggaa ggttccggcc tgcgtatcg 1980
gctggatatg attgccaccc tcctctccaa cggatcgctc gttgccaag tgacccaggaa 2040
aaacagcgat gaatatggcg ttcgcagat cttcatcgat attgaagtgg ataagctgat 2100
cgacggcgca acccgcgacg ccaagctgca acggattatg gatttcatca ccaccgcga 2160
gcgcgcgcgt gaaaatgtgg cggtcgtct tcctggccat gaatttaccc gtcgtcgaa 2220
tgaaaaccgc cgcacggca ttaccgtcga tgacagcgta tggccaaaaa ttcaaggcgct 2280
gtaaggagct cacccatgac agcgtatggg caaaaattca ggcgtgtaa ggagctcacc 2340
catgatttt ggtcatattt ctcaaccaa tccgtgtcgat ctgcccgcgg ccattgagcg 2400
ggcgttgcgtat ttccgtcgc cgcaggattt ccacgcgtcg gcacccggcg tcgtggaaat 2460
cgacggccaa aacatcttcg cgcaggatccatcgacttacc acgtcgatcg ccgtgtaaaaa 2520
tcgtccggag gtccaccgtc gctatctggat tatccagttt ctggcatcg ggcgaaaaa 2580
aatcggatc gccattgata cggcaataa tcaaatcagc gaatctttat tagaacagcg 2640
cgatattatt ttttatcagc acagcgaaca tgaatcgatc tttgaaatga cggcaggca 2700
ctatgcgata ttccgcgtcc aagatgttca tcgtcctggat tggataaaaaa ctgtagccac 2760
ggcgttgcgtat ccgcgtccgc aaaaatgtcg ttaaagtcgc tatttcagtt ttataagaag ggcacaaaaa 2820

tgaattcgaa taataccggc tacattatcg gtgcgtaccc ctgtcccccc tgcgcaccct 2880
 catttcacca aaagagtgaa gaggaagaga tggaaattctg gcggcagctc tccgacaccc 2940
 cggatattcg cgggctggag caaccctgcc taccctgcct tgaacatctt catccgctcg 3000
 gcgacgagtg gttattgcgc catacccccgg gacactggca gattgtcggt accgcccattca 3060
 tggaaaccat gcccgcgc ggtaaaaacg gcggcttgg gctggcgtcc agcgacgaaa 3120
 cgtagcgcggaa agcctgcgtg gaggactatc gccacctgcgca gcagaagatc gctaaaatca 3180
 atggcaatac cggccggaaag gtcattgccc ttgagcttca cggcccccgg ctggcgggca 3240
 atgccaacgt ggctcaggct accgcgcct ttgcccgttc attaaaagaa attaccgct 3300
 gggactggc tcgctggagc actgcgcacgc gatgaccggc agcgccgcgc 3360
 gcaaggatt tttgccgtta gaaaacgtgc tggaaagccat tgccgattat gacgttggca 3420
 tttgtattaa ctggcgcgt tcggccattt aaggcggaa taccgtgcta ccgctcaccc 3480
 atacgcagca ggtaaaacgg gcaggaaagc tcggcgcgt gatgtttctt ggcacgcacgc 3540
 agaccggcga gtacggcga tggcaggatt tacacgcgc gttcgcgcct ttctgcccgc 3600
 agagcctgtat gaccacccaa cacgcctgtg aattatttgc ctgcgcagga accggcccccc 3660
 tgcaattttc aggcataaaa ttactggaaa ttaatgcggc cggaaacgtt gatcatgcgca 3720
 tcgcgatatt acgcgcacggc atctccgcgc taaaacaagg acaataataa taatcacctt 3780
 catcaccaga atattttaa tattacgaga ctataaagat gaatataacc tctaactcta 3840
 caaccaaaga tataccgcgc cagcgttgtt taagaatcat tccgcctata ctgatcacctt 3900
 gtattatttc ttatatggac cgggtcaata ttgccttgc gatgcccggg ggtatggatg 3960
 ccgacttagg tatttccgc accatggcgg ggctggcggg cggtagcggt aagaattta 4020
 atctattttt acaggttccc ggcggggaaa ttgcccgttca cggtagcggt aagaattta 4080
 tcgctggc tcgttgtcgc tggcgggtca tctccgtgtt gacgggggtt attaccaatc 4140
 agtaccagct gctggccctg cgcttcttac tggcgtggc ggaaggcggt atgctggcg 4200
 tcgttctcac gatgatcagt aactggttcc ccgacgcgtga acgcggcgc gccaacgcga 4260
 ttgtcattat gtttgtgcgcg attgccgggaa ttatcaccgc cccactctca ggctggatta 4320
 tcacggttct cgtactggcgc tggctgttta ttatcgaagg tttgcctctcg ctggttgttc 4380
 tgggtctgtg ggcatacacc atctatgacc gtccgcgggaa agcgcgcgtgg atttccgaag 4440
 cagagaagcg ctatctggc gagacgcgtgg ccgcggggaa aaaagccatt gccggcaccg 4500
 aggtgaaaaaa cgcctctcg agcgcgcgttc tctccgacaa aaccatgtgg cagcttatcg 4560
 ccctgaacctt cttctaccag accgcattt acggctacac cctgtggctt cccaccattc 4620
 tggaaattt gacccatagc agcatggggc aggtcggtat gcttgcatt ctggcgtacg 4680
 tcggcgcatt tgctggatg ttccctttt cctcccttcc agaccgaacc ggtaaacgcga 4740
 agctgttctgt ctgcctggc ctgattggct tgcgtctgtt catgttccctg tcggcggcgc 4800
 tggaaattt gacccatagc agcatggggc aggtcggtat gcttgcatt ctggcgtacg 4860
 cggcggcgtgg cgtgttctgg accatcccgg cacgtctgtt cagcgcggaa atggcggggcg 4920
 gcgcgcgcgg ggttatcaac gcgcgtggca acctcgccggg attttgcgcg ccttatgcgg 4980
 tcgggggtat gatcacgtt tacagcaaag acgctggcgat ctattgcctt gcatctccc 5040
 tggcgtggc cgcgtgtat gcgcgtgtc tgccggcgaat atgcgtatgc ggtgtgcgc 5100
 cggtaaaagac gataaaatcca cataaaacgcata ctcgtaaac tcgagccggc cggcgtgcg 5160
 cctgcggggc ctgcgaaata tgccgggttcc acccggttaac aatgagatgc gaaagatgag 5220
 caagaaacag gccttctggc tgggtattga ttgcggcggc acctatctga aagccggttt 5280
 atatgacgcgc gaaaggcatg aacatggcat tgcgtggca ggcgtacggc cgtatgtcgcc 5340
 cctgcgggtt tacgcccgaac gcgcacatgcg ccagctctgg caacactgcg cggcgcaccat 5400
 tggcggcgttacatcagcagg cagggttatac cggcgtggca acctatctga aagccggttt 5460
 cgcgtgggtt caagggtctt ttctcccttca taagcaggat cggccgtgg gtaacgcatt 5520
 cctctcccttcc gatcgtcggtt cgtgtttttt cgttgcgttcc gtcgtggcc accgtattcc 5580
 cgaacggcgtt tatcccttta cccggcagac gtcgtggacc ggacatccggg cttctttgtt 5640
 cgcgtgggtt aaagagaatg aaccccgacgc ctacgcgcgaa attggctgcg tgcgtatggg 5700

gcatgactat ctgcgctggc gcttaaccgg cgCGaaggGC tgcgaggaga gcaacatctc 5760
 cgagtccaaC ctctacaaca tggccatggc ccagtacgac ccgcgcctga ccgagtggct 5820
 gggcatcggt gaaatcgata ggcgcgtgc ccccgttGta gggtcagccg aaatttgcgg 5880
 ggagatcacc gtcaggcag ccgcttaac cggctcgccg gcccgtactc ccgtcggtgg 5940
 cggcctgttt gacgtggctc ccaccgcct ttgcgcccggg attgaggatg agtcgaccct 6000
 caatcggtg atggggaccc gggccgtac tagcggtatac gtcacggcc tgcgacca 6060
 tgagccccac ctttacgtct atggccgcta cgtcaatgac ggccagtata tcgttacgca 6120
 agccagcccc acctcatccg gcaacctcga atggtttacc gcccagtggg gcgatctctc 6180
 gtttgcgtg atcaatcagg cgcgtccag cctgcccggaa gcccggagcg agctgtttt 6240
 tctgccgtt ctgtatggca gcaacgcgg gctggagatg acctgcggct tttacggcat 6300
 gcaggcgtg catacccgcg cgcacctgtc gcaggcgggt tatgaaggcg tggtagtttag 6360
 ccatatgacc cacctcagcc gtatgcgcga acgcatttaca aacgttccagg ccctgcgcgt 6420
 caccggcgcc cccggcgcact ccgcacgtctg gatgcagatg ctggcggacg taagcggctt 6480
 acgcattgaa ctccccgaagg tggaaagagac cgcgtgtttt ggccggggcc tcggcgtcg 6540
 tgtcggtacc ggcgtatacc gcagctttag cgaagccccgg cgcgcggccg agcaccgggt 6600
 ggcgcacgtg ctgcccata tgaccggcca cgcgcgtat cagcgcataat accgcccata 6660
 cctgcatttg attgaagcac tacaggcata tcacgcccgt attaaggagc acgcattatg 6720
 agccgaccat tactgcagct ggcgcgtcgac cataccagcc ttcaaggctgc gcagcgcgt 6780
 gtcgcctgc tacaggatca cgttgcattt gttggaggcgg gaaccatct ctgtttaacc 6840
 gaagggctta ggcgcgttaa agccctgcgc gcccagtgtc cggggaaagat catcgctgcc 6900
 gactggaaag tcgcccacgc cggtaaaacc ctggcgcagc aggcctttgg cgctggcgcc 6960
 aactggatga ccatcatttg cgcgcacccg ctgcacccgg tcgagaaaagg ccacgcgtg 7020
 gcccaggcct gggcggtga aattcagatg gagctgttcg gcaactggac gctggatgac 7080
 gcccgcgcct ggtaccgtac cggcgtccat caggcgtattt accatcgccg acgcgtgc 7140
 caggccagcg ggcagcgtg gggggaggcg gatctggcgc gcatgaaagc gctgtccgt 7200
 attggccttgc agctatcgat taccggcgc attaccccg cgcgttacc gctgttcaaa 7260
 gatatcaacg tcaaagcctt tattggccgg cgcgcgtcg caggcgcgc ccattccggcg 7320
 cgggttgcgc cggaaattcca cgcgcataatc gacgctatct ggggagaaca gcatgcgtaa 7380
 ccacccgtta ggtatTTatg aaaaagcgt ggcgaaggat ctcagctggc ctgagcggct 7440
 ggtactggcc aaaagctgcg gtttgattt tgcgaaatg tcgggtggacg agaccgatga 7500
 acgccttcg cgcctggagt ggaccccgcc ccagcgcgc tgcgtgtga ggcgcgtct 7560
 gggaaaccgcg gtcgcatttcc ctcgtatgtc ttgtccgc catgcgcgtt tccctttgg 7620
 cagccgcgtat gaaagggtac ggcgttgcg ggcggatgatt atgaccaaaag ccattccgc 7680
 ggcgcgcgtat ctggggatcc gcaccatcca gtcggcggt tacgacgtct attacgaaga 7740
 gcatgatgaa ggcacccggc agcgtttgc cgaagggtcg gcctggcgg tagaacaggc 7800
 cgcgcgcgtat cggtaatgc tggcggtgg gatcatggac accgcctta tgaactccat 7860
 cagcaaatgg aaaaagtgg acgagatgt ttgcgtcaccg tggtttaccg tctacccgg 7920
 cgtcgcaac ctcagcgctt gggaaacga cgtcaccgcg gagctgaaagc tggcgtcg 7980
 tcgtatcgcc gccatccacc taaaagatac gtcggcggt accgcgtatc gcccggcca 8040
 gttcccgac gtcggcggt gcaaggatg cgtcgattt gtcggcattt ttaagacgct 8100
 ggcgcgtcg aactaccgcg gttcattttt gattgagatg tggacggaga aagccagcga 8160
 gcccggcggt gggatattcc aggccggcg ctggatcgaa tcacggatgc aggaaggggg 8220
 attcacatgt tagaacaact gaaagccgag gtactggcgg caaacctggc cctcccccgc 8280
 cacggcctgg tcaccttac ctggggcaac gtcagcgcgg tcgatgaaac ggcgcgtcg 8340
 atggtcatta agccttccgg cgtcgaatatac gaggatgt gtcggcgtcg tttttttttt 8400
 gtagagatgg ccagcggtaa agtcgttgc ggcggtaaaa aacccttcc agatacgcca 8460
 acgcatttgcg cgcgttatcc cgcgtatcc cagatcgccg ggatcgacca taccactcc 8520
 cgcgcgcgtat cgcgttgcg cgcgttgcg cgcgttgcg cgcgttgcg cgcgttgcg 8580

gccgactact tctatggcgc gatccctgt acccgacgga tgaccgttga ggagattaac 8640
 ggcgagttatg agtatacgac cggcgagggtt attatcaaaa cctttaaca ggcgggcctg 8700
 gatccggcgc aaatcccggc ggtattggtc cattcacacg gcccctttgc ctggggtaaa 8760
 gacgcccggc acgcccgtaca taaccccgtg gtgctggagg agtgcgccta catgggcctc 8820
 ttctcgcgcc agtgccaca gctgccggat atgcagtctg aactgctcgtaaaacactat 8880
 ctgcgtaaac acggcgccaa cgctattacg ggcaaaaacta gtcccgccga actccccgga 8940
 taaggcgctt tggcccccgg gggaaagcgtg caggatgttgc ctgaactttc ccggagcgat 9000
 gctgcgcatac tgtccggct acgcgtcccc ggcgcctgc ggtcagcacc gcgcggcg 9060
 gaaaacccat caaccctacg ccgaatataat atgtccttgc agtaacgacg cttccacgcc 9120
 gccggtccag gctgggtgtc ttgcgaaaa tcttgcgaaa atagccgaca tcgttaaacc 9180
 cgcatattcat cgccacctcg gtaatcgaca gggaaatcgct gataagcagc tttccggcg 9240
 cccttacccg ctgacgggtgc agcgcttcgg taacgtcagc cgaaagcat ggcgataaac 9300
 ggccccagat aaccggcggtt gcagtgacgc tcct 9334

<210> 20

<211> 282

<212> PRT

<213> YiaJ-Ec

<400> 20

Met Gly Lys Glu Val Met Gly Lys Lys Glu Asn Glu Met Ala Gln Glu

1 5 10 15

Lys Glu Arg Pro Ala Gly Ser Gln Ser Leu Phe Arg Gly Leu Met Leu

20 25 30

Ile Glu Ile Leu Ser Asn Tyr Pro Asn Gly Cys Pro Leu Ala His Leu

35 40 45

Ser Glu Leu Ala Gly Leu Asn Lys Ser Thr Val His Arg Leu Leu Gln

50 55 60

Gly Leu Gln Ser Cys Gly Tyr Val Thr Thr Ala Pro Ala Ala Gly Ser

65 70 75 80

Tyr Arg Leu Thr Thr Lys Phe Ile Ala Val Gly Gln Lys Ala Leu Ser

85 90 95

Ser Leu Asn Ile Ile His Ile Ala Ala Pro His Leu Glu Ala Leu Asn

100 105 110

Ile Ala Thr Gly Glu Thr Ile Asn Phe Ser Ser Arg Glu Asp Asp His

115 120 125

Ala Ile Leu Ile Tyr Lys Leu Glu Pro Thr Thr Gly Met Leu Arg Thr

130 135 140

Arg Ala Tyr Ile Gly Gln His Met Pro Leu Tyr Cys Ser Ala Met Gly

145	150	155	160
-----	-----	-----	-----

Lys Ile Tyr Met Ala Phe Gly His Pro Asp Tyr Val Lys Ser Tyr Trp			
165	170	175	

Glu Ser His Gln His Glu Ile Gln Pro Leu Thr Arg Asn Thr Ile Thr			
180	185	190	

Glu Leu Pro Ala Met Phe Asp Glu Leu Ala His Ile Arg Glu Ser Gly			
195	200	205	

Ala Ala Met Asp Arg Glu Glu Asn Glu Leu Gly Val Ser Cys Ile Ala			
210	215	220	

Val Pro Val Phe Asp Ile His Gly Arg Val Pro Tyr Ala Val Ser Ile			
225	230	235	240

Ser Leu Ser Thr Ser Arg Leu Lys Gln Val Gly Glu Lys Asn Leu Leu			
245	250	255	

Lys Pro Leu Arg Glu Thr Ala Gln Ala Ile Ser Asn Glu Leu Gly Phe			
260	265	270	

Thr Val Arg Asp Asp Leu Gly Ala Ile Thr			
275	280		

<210> 21

<211> 268

<212> PRT

<213> YiaJ-Hi

<400> 21

Met Asn Ile Glu Val Lys Met Glu Lys Glu Lys Ser Leu Gly Asn Gln			
1	5	10	15

Ala Leu Ile Arg Gly Leu Arg Leu Leu Asp Ile Leu Ser Asn Tyr Pro			
20	25	30	

Asn Gly Cys Pro Leu Ala Lys Leu Ala Glu Leu Ala Asn Leu Asn Lys			
35	40	45	

Ser Thr Ala His Arg Leu Leu Gln Gly Leu Gln Asn Glu Gly Tyr Val			
50	55	60	

Lys Pro Ala Asn Ala Ala Gly Ser Tyr Arg Leu Thr Ile Lys Cys Leu			
65	70	75	80

Ser Ile Gly Gln Lys Val Leu Ser Ser Met Asn Ile Ile His Val Ala
85 90 95

Ser Pro Tyr Leu Glu Gln Leu Asn Leu Lys Leu Gly Glu Thr Ile Asn
 100 105 110

Phe Ser Lys Arg Glu Asp Asp His Ala Ile Met Ile Tyr Lys Leu Glu
115 120 125

Pro Thr Asn Gly Met Leu Lys Thr Arg Ala Tyr Ile Gly Gln Tyr Leu
130 135 140

Lys Leu Tyr Cys Ser Ala Met Gly Lys Ile Phe Leu Ala Tyr Glu Lys
 145 150 155 160

Lys Val Asp Tyr Leu Ser His Tyr Trp Gln Ser His Gln Arg Glu Ile
 165 170 175

Lys Lys Leu Thr Arg Tyr Thr Ile Thr Glu Leu Asp Asp Ile Lys Leu
180 185 190

Glu Leu Glu Thr Ile Arg Gln Thr Ala Tyr Ala Met Asp Arg Glu Glu
..... 195 200 205

Asn Glu Leu Gly Val Thr Cys Ile Ala Cys Pro Ile Phe Asp Ser Phe
210 215 220

Gly Gln Val Glu Tyr Ala Ile Ser Val Ser Met Ser Ile Tyr Arg Leu
225 . . . 230 . . . 235 . . . 240

Asn Lys Phe Gly Thr Asp Ala Phe Leu Gln Glu Ile Arg Lys Thr Ala
245 250 255

Glu Gln Ile Ser Leu Glu Leu Gly Tyr Glu Asn Ile
260 265

<210> 22

<211> 332

<212> PRT

<213> YiaK-EG

<400> 22

Met Lys Val Thr Phe Glu Gln Leu Lys Ala Ala Phe Asn Arg Val Leu
1 5 10 15

Ile Ser Arg Gly Val Asp Ser Glu Thr Ala Asp Ala Cys Ala Glu Met
20 25 30

Phe Ala Arg Thr Thr Glu Ser Gly Val Tyr Ser His Gly Val Asn Arg
 35 40 45

Phe Pro Arg Phe Ile Gln Gln Leu Glu Asn Gly Asp Ile Ile Pro Asp
 50 55 60

Ala Gln Pro Lys Arg Ile Thr Ser Leu Gly Ala Ile Glu Gln Trp Asp
 65 70 75 80

Ala Gln Arg Ser Ile Gly Asn Leu Thr Ala Lys Lys Met Met Asp Arg
 85 90 95

Ala Ile Glu Leu Ala Ala Asp His Gly Ile Gly Leu Val Ala Leu Arg
 100 105 110

Asn Ala Asn His Trp Met Arg Gly Gly Ser Tyr Gly Trp Gln Ala Ala
 115 120 125

Glu Lys Gly Tyr Ile Gly Ile Cys Trp Thr Asn Ser Ile Ala Val Met
 130 135 140

Pro Pro Trp Gly Ala Lys Glu Cys Arg Ile Gly Thr Asn Pro Leu Ile
 145 150 155 160

Val Ala Ile Pro Ser Thr Pro Ile Thr Met Val Asp Met Ser Met Ser
 165 170 175

Met Phe Ser Tyr Gly Met Leu Glu Val Asn Arg Leu Ala Gly Arg Gln
 180 185 190

Leu Pro Val Asp Gly Gly Phe Asp Asp Glu Gly Asn Leu Thr Lys Glu
 195 200 205

Pro Gly Val Ile Glu Lys Asn Arg Arg Ile Leu Pro Met Gly Tyr Trp
 210 215 220

Lys Gly Ser Gly Met Ser Ile Val Leu Asp Met Ile Ala Thr Leu Leu
 225 230 235 240

Ser Asp Gly Ala Ser Val Ala Glu Val Thr Gin Asp Asn Ser Asp Glu
 245 250 255

Tyr Gly Ile Ser Gln Ile Phe Ile Ala Ile Glu Val Asp Lys Leu Ile
 260 265 270

Asp Gly Pro Thr Arg Asp Ala Lys Leu Gln Arg Ile Met Asp Tyr Val
 275 280 285

Thr Ser Ala Glu Arg Ala Asp Glu Asn Gln Ala Ile Arg Leu Pro Gly
 290 295 300

His Glu Phe Thr Thr Leu Leu Ala Glu Asn Arg Arg Asn Gly Ile Thr
 305 310 315 320

Val Asp Asp Ser Val Trp Ala Lys Ile Gln Ala Leu
 325 330

<210> 23

<211> 332

<212> PRT

<213> YiaK-Hi

<400> 23

Met Arg Val Ser Tyr Asp Glu Leu Lys Asn Glu Phe Lys Arg Val Leu
 1 5 10 15

Leu Asp Arg Gln Leu Thr Glu Glu Leu Ala Glu Glu Cys Ala Thr Ala
 20 25 30

Phe Thr Asp Thr Thr Gln Ala Gly Ala Tyr Ser His Gly Ile Asn Arg
 35 40 45

Phe Pro Arg Phe Ile Gln Gln Leu Glu Gln Gly Asp Ile Val Pro Asn
 50 55 60

Ala Ile Pro Thr Lys Val Leu Ser Leu Gly Ser Ile Glu Gln Trp Asp
 65 70 75 80

Ala His Gln Ala Ile Gly Asn Leu Thr Ala Lys Lys Met Met Asp Arg
 85 90 95

Ala Ile Glu Leu Ala Ser Gln His Gly Val Gly Val Ile Ala Leu Arg
 100 105 110

Asn Ala Asn His Trp Met Arg Gly Gly Ser Tyr Gly Trp Gln Ala Ala
 115 120 125

Glu Lys Gly Tyr Ile Gly Ile Cys Trp Thr Asn Ala Leu Ala Val Met
 130 135 140

Pro Pro Trp Gly Ala Lys Glu Cys Arg Ile Gly Thr Asn Pro Leu Ile
 145 150 155 160

Ile Ala Val Pro Thr Thr Pro Ile Thr Met Val Asp Met Ser Cys Ser

165

170

175

Met Tyr Ser Tyr Gly Met Leu Glu Val His Arg Leu Ala Gly Arg Gln
 180 185 190

Thr Phe Val Asp Ala Gly Phe Asp Asp Glu Gly Asn Leu Thr Arg Asp
 195 200 205

Pro Ser Ile Val Glu Lys Asn Arg Arg Leu Leu Pro Met Gly Phe Trp
 210 215 220

Lys Gly Ser Gly Leu Ser Ile Val Leu Asp Met Ile Ala Thr Leu Leu
 225 230 235 240

Ser Asn Gly Glu Ser Thr Val Ala Val Thr Glu Asp Lys Asn Asp Glu
 245 250 255

Tyr Cys Val Ser Gln Val Phe Ile Ala Ile Glu Val Asp Arg Leu Ile
 260 265 270

Asp Gly Lys Ser Lys Asp Glu Lys Leu Asn Arg Ile Met Asp Tyr Val
 275 280 285

Lys Thr Ala Glu Arg Ser Asp Pro Thr Gln Ala Val Arg Leu Pro Gly
 290 295 300

His Glu Phe Thr Thr Ile Leu Ser Asp Asn Gln Thr Asn Gly Ile Pro
 305 310 315 320

Val Asp Glu Arg Val Trp Ala Lys Leu Lys Thr Leu
 325 330

<210> 24

<211> 155

<212> PRT

<213> YiaL-Ec

<400> 24

Met Ile Phe Gly His Ile Ala Gln Pro Asn Pro Cys Arg Leu Pro Ala
 1 5 10 15

Ala Ile Glu Lys Ala Leu Asp Phe Leu Arg Ala Thr Asp Phe Asn Ala
 20 25 30

Leu Glu Pro Gly Val Val Glu Ile Asp Gly Lys Asn Ile Tyr Thr Gln
 35 40 45

Ile Ile Asp Leu Thr Thr Arg Glu Ala Val Val Asn Arg Pro Glu Val
 50 55 60

His Arg Arg Tyr Ile Asp Ile Gln Phe Leu Ala Trp Gly Glu Glu Lys
 65 70 75 80

Ile Gly Ile Ala Ile Asp Thr Gly Asn Asn Lys Val Ser Glu Ser Leu
 85 90 95

Leu Glu Gln Arg Asn Ile Ile Phe Tyr His Asp Ser Glu His Glu Ser
 100 105 110

Phe Ile Glu Met Ile Pro Gly Ser Tyr Ala Ile Phe Phe Pro Gln Asp
 115 120 125

Val His Arg Pro Gly Cys Ile Met Gln Thr Ala Ser Glu Ile Arg Lys
 130 135 140

Ile Val Val Lys Val Ala Leu Thr Ala Leu Asn
 145 150 155

<210> 25

<211> 155

<212> PRT

<213> YiaL-Hi

<400> 25

Met Ile Ile Ser Ser Leu Thr Asn Pro Asn Phe Lys Val Gly Leu Pro
 1 5 10 15

Lys Val Ile Ala Glu Val Cys Asp Tyr Leu Asn Thr Leu Asp Leu Asn
 20 25 30

Ala Leu Glu Asn Gly Arg His Asp Ile Asn Asp Gln Ile Tyr Met Asn
 35 40 45

Val Met Glu Pro Glu Thr Ala Glu Pro Ser Ser Lys Lys Ala Glu Leu
 50 55 60

His His Glu Tyr Leu Asp Val Gln Val Leu Ile Arg Gly Thr Glu Asn
 65 70 75 80

Ile Glu Val Gly Ala Thr Tyr Pro Asn Leu Ser Lys Tyr Glu Asp Tyr
 85 90 95

Asn Glu Ala Asp Asp Tyr Gln Leu Cys Ala Asp Ile Asp Asp Lys Phe
 100 105 110

Thr Val Thr Met Lys Pro Lys Met Phe Ala Val Phe Tyr Pro Tyr Glu
 115 120 125

Pro His Lys Pro Cys Cys Val Val Asn Gly Lys Thr Glu Lys Ile Lys
 130 135 140

Lys Leu Val Val Lys Val Pro Val Lys Leu Ile
 145 150 155

<210> 26

<211> 498

<212> PRT

<213> LyxK-Ec

<400> 26

Met Thr Gln Tyr Trp Leu Gly Leu Asp Cys Gly Gly Ser Trp Leu Lys
 1 5 10 15

Ala Gly Leu Tyr Asp Arg Glu Gly Arg Glu Ala Gly Val Gln Arg Leu
 20 25 30

Pro Leu Cys Ala Leu Ser Pro Gln Pro Gly Trp Ala Glu Arg Asp Met
 35 40 45

Ala Glu Leu Trp Gln Cys Cys Met Ala Val Ile Arg Ala Leu Leu Thr
 50 55 60

His Ser Gly Val Ser Gly Glu Gln Ile Val Gly Ile Gly Ile Ser Ala
 65 70 75 80

Gln Gly Lys Gly Leu Phe Leu Leu Asp Lys Asn Asp Lys Pro Leu Gly
 85 90 95

Asn Ala Ile Leu Ser Ser Asp Arg Arg Ala Met Glu Ile Val Arg Arg
 100 105 110

Trp Gln Glu Asp Gly Ile Pro Glu Lys Leu Tyr Pro Leu Thr Arg Gln
 115 120 125

Thr Leu Trp Thr Gly His Pro Val Ser Leu Leu Arg Trp Leu Lys Glu
 130 135 140

His Glu Pro Glu Arg Tyr Ala Gln Ile Gly Cys Val Met Met Thr His
 145 150 155 160

Asp Tyr Leu Arg Trp Cys Leu Thr Gly Val Lys Gly Cys Glu Glu Ser

165	170	175
Asn Ile Ser Glu Ser Asn Leu Tyr Asn Met Ser Leu Gly Glu Tyr Asp		
180	185	190
Pro Cys Leu Thr Asp Trp Leu Gly Ile Ala Glu Ile Asn His Ala Leu		
195	200	205
Pro Pro Val Val Gly Ser Ala Glu Ile Cys Gly Glu Ile Thr Ala Gln		
210	215	220
Thr Ala Ala Leu Thr Gly Leu Lys Ala Gly Thr Pro Val Val Gly Gly		
225	230	235
Leu Phe Asp Val Val Ser Thr Ala Leu Cys Ala Gly Ile Glu Asp Glu		
245	250	255
Phe Thr Leu Asn Ala Val Met Gly Thr Trp Ala Val Thr Ser Gly Ile		
260	265	270
Thr Arg Gly Leu Arg Asp Gly Glu Ala His Pro Tyr Val Tyr Gly Arg		
275	280	285
Tyr Val Asn Asp Gly Glu Phe Ile Val His Glu Ala Ser Pro Thr Ser		
290	295	300
Ser Gly Asn Leu Glu Trp Phe Thr Ala Gln Trp Gly Glu Ile Ser Phe		
305	310	315
Asp Glu Ile Asn Gln Ala Val Ala Ser Leu Pro Lys Ala Gly Gly Asp		
325	330	335
Leu Phe Phe Leu Pro Phe Leu Tyr Gly Ser Asn Ala Gly Leu Glu Met		
340	345	350
Thr Ser Gly Phe Tyr Gly Met Gln Ala Ile His Thr Arg Ala His Leu		
355	360	365
Leu Gln Ala Ile Tyr Glu Gly Val Val Phe Ser His Met Thr His Leu		
370	375	380
Asn Arg Met Arg Glu Arg Phe Thr Asp Val His Thr Leu Arg Val Thr		
385	390	395
Gly Gly Pro Ala His Ser Asp Val Trp Met Gln Met Leu Ala Asp Val		
405	410	415
Ser Gly Leu Arg Ile Glu Leu Pro Gln Val Glu Glu Thr Gly Cys Phe		

420 425 430

Gly Ala Ala Leu Ala Ala Arg Val Gly Thr Gly Val Tyr His Asn Phe
435 440 445

Ser Glu Ala Gln Arg Asp Leu Arg His Pro Val Arg Thr Leu Leu Pro
450 455 460

Asp Met Thr Ala His Gln Leu Tyr Gln Lys Lys Tyr Gln Arg Tyr Gln
465 470 475 480

His Leu Ile Ala Ala Leu Gln Gly Phe His Ala Arg Ile Lys Glu His
485 490 495

Thr Leu

<210> 27

<211> 485

<212> PRT

<213> LyxK-Hi

<400> 27

Met His Tyr Tyr Leu Gly Ile Asp Cys Gly Gly Thr Phe Ile Lys Ala
1 5 10 15

Ala Ile Phe Asp Gln Asn Gly Thr Leu Gln Ser Ile Ala Arg Arg Asn
20 25 30

Ile Pro Ile Ile Ser Glu Lys Pro Gly Tyr Ala Glu Arg Asp Met Asp
35 40 45

Glu Leu Trp Asn Leu Cys Ala Gln Val Ile Gln Lys Thr Ile Arg Gln
50 55 60

Ser Ser Ile Leu Pro Gln Gln Ile Lys Ala Ile Gly Ile Ser Ala Gln
65 70 75 80

Gly Lys Gly Ala Phe Phe Leu Asp Lys Asp Asn Lys Pro Leu Gly Arg
85 90 95

Ala Ile Leu Ser Ser Asp Gln Arg Ala Tyr Glu Ile Val Gln Cys Trp
100 105 110

Gln Lys Glu Asn Ile Leu Gln Lys Phe Tyr Pro Ile Thr Leu Gln Thr
115 120 125

Leu Trp Met Gly His Pro Val Ser Ile Leu Arg Trp Ile Lys Glu Asn
130 135 140

Glu Pro Ser Arg Tyr Glu Gln Ile His Thr Ile Leu Met Ser His Asp
145 150 155 160

Tyr Leu Arg Phe Cys Leu Thr Glu Lys Leu Tyr Cys Glu Glu Thr Asn
165 170 175

Ile Ser Glu Ser Asn Phe Tyr Asn Met Arg Glu Gly Lys Tyr Asp Ile
180 185 190

Gln Leu Ala Lys Leu Phe Gly Ile Thr Glu Cys Ile Asp Lys Leu Pro
195 200 205

Pro Ile Ile Lys Ser Asn Lys Ile Ala Gly Tyr Val Thr Ser Arg Ala
210 215 220

Ala Glu Gln Ser Gly Leu Val Glu Gly Ile Pro Val Val Gly Gly Leu
225 230 235 240

Phe Asp Val Val Ser Thr Ala Leu Cys Ala Asp Leu Lys Asp Asp Gln
245 250 255

His Leu Asn Val Val Leu Gly Thr Trp Ser Val Val Ser Gly Val Thr
260 265 270

His Tyr Ile Asp Asp Asn Gln Thr Ile Pro Phe Val Tyr Gly Lys Tyr
275 280 285

Pro Glu Lys Asn Lys Phe Ile Ile His Glu Ala Ser Pro Thr Ser Ala
290 295 300

Gly Asn Leu Glu Trp Phe Val Asn Gln Phe Asn Leu Pro Asn Tyr Asp
305 310 315 320

Asp Ile Asn His Glu Ile Ala Lys Leu Lys Pro Ala Ser Ser Val
325 330 335

Leu Phe Ala Pro Phe Leu Tyr Gly Ser Asn Ala Lys Leu Gly Met Gln
340 345 350

Ala Gly Phe Tyr Gly Ile Gln Ser His His Thr Gln Ile His Leu Leu
355 360 365

Gln Ala Ile Tyr Glu Gly Val Ile Phe Ser Leu Met Ser His Leu Glu
370 375 380

Arg Met Gln Val Arg Phe Pro Asn Ala Ser Thr Val Arg Val Thr Gly
 385 390 395 400

Gly Pro Ala Lys Ser Glu Val Trp Met Gln Met Leu Ala Asp Ile Ser
 405 410 415

Gly Met Arg Leu Glu Ile Pro Asn Ile Glu Glu Thr Gly Cys Leu Gly
 420 425 430

Ala Ala Leu Met Ala Met Gln Ala Glu Ser Ala Val Glu Ile Ser Gln
 435 440 445

Ile Leu Asn Ile Asp Arg Lys Ile Phe Leu Pro Asp Lys Asn Gln Tyr
 450 455 460

Ser Lys Tyr Gln His Lys Tyr His Arg Tyr Leu Lys Phe Ile Glu Ala
 465 470 475 480

Leu Lys Asn Leu Asp
 485

<210> 28

<211> 220

<212> PRT

<213> YiaQ-Ec

<400> 28

Met Ser Arg Pro Leu Leu Gln Leu Ala Leu Asp His Ser Ser Leu Glu
 1 5 10 15

Ala Ala Gln Arg Asp Val Thr Leu Leu Lys Asp Ser Val Asp Ile Val
 20 25 30

Glu Ala Gly Thr Ile Leu Cys Leu Asn Glu Gly Leu Gly Ala Val Lys
 35 40 45

Ala Leu Arg Glu Gln Cys Pro Asp Lys Ile Ile Val Ala Asp Trp Lys
 50 55 60

Val Ala Asp Ala Gly Glu Thr Leu Ala Gln Gln Ala Phe Gly Ala Gly
 65 70 75 80

Ala Asn Trp Met Thr Ile Ile Cys Ala Ala Pro Leu Ala Thr Val Glu
 85 90 95

Lys Gly His Ala Met Ala Gln Arg Cys Gly Gly Glu Ile Gln Ile Glu
 100 105 110

Leu Phe Gly Asn Trp Thr Leu Asp Asp Ala Arg Asp Trp His Arg Ile
 115 120 125

Gly Val Arg Gln Ala Ile Tyr His Arg Gly Arg Asp Ala Gln Ala Ser
 130 135 140

Gly Gln Gln Trp Gly Glu Ala Asp Leu Ala Arg Met Lys Ala Leu Ser
 145 150 155 160

Asp Ile Gly Leu Glu Leu Ser Ile Thr Gly Gly Ile Thr Pro Ala Asp
 165 170 175

Leu Pro Leu Phe Lys Asp Ile Arg Val Lys Ala Phe Ile Ala Gly Arg
 180 185 190

Ala Leu Ala Gly Ala Ala Asn Pro Ala Gln Val Ala Gly Asp Phe His
 195 200 205

Ala Gln Ile Asp Ala Ile Trp Gly Gly Ala Arg Ala
 210 215 220

<210> 29

<211> 225

<212> PRT

<213> YiaQ-Hi

<400> 29

Met Gly Lys Pro Leu Leu Gln Ile Ala Leu Asp Ala Gln Tyr Leu Glu
 1 5 10 15

Thr Ala Leu Val Asp Val Lys Gln Ile Glu His Asn Ile Asp Ile Ile
 20 25 30

Glu Val Gly Thr Ile Leu Ala Cys Ser Glu Gly Met Arg Ala Val Arg
 35 40 45

Ile Leu Arg Ala Leu Tyr Pro Asn Gln Ile Leu Val Cys Asp Leu Lys
 50 55 60

Thr Thr Asp Ala Gly Ala Thr Leu Ala Lys Met Ala Phe Glu Ala Gly
 65 70 75 80

Ala Asp Trp Leu Thr Val Ser Ala Ala Ala His Pro Ala Thr Lys Ala
 85 90 95

Ala Cys Gln Lys Val Ala Glu Glu Phe Asn Lys Ile Gln Pro Asn Leu

100

105

110

Gly Val Pro Lys Glu Ile Gln Ile Glu Leu Tyr Gly Asn Trp Asn Phe
 115 120 125

Asp Glu Val Lys Asn Trp Leu Gln Leu Gly Ile Lys Gln Ala Ile Tyr
 130 135 140

His Arg Ser Arg Asp Ala Glu Leu Ser Gly Leu Ser Trp Ser Asn Gln
 145 150 155 160

Asp Ile Glu Asn Ile Glu Lys Leu Asp Ser Leu Gly Ile Glu Leu Ser
 165 170 175

Ile Thr Gly Gly Ile Thr Pro Asp Asp Leu His Leu Phe Lys Asn Thr
 180 185 190

Lys Asn Leu Lys Ala Phe Ile Ala Gly Arg Ala Leu Val Gly Lys Ser
 195 200 205

Gly Arg Glu Ile Ala Glu Gln Leu Lys Gln Lys Ile Gly Gln Phe Trp
 210 215 220

Ile
 225

<210> 30
 <211> 297
 <212> PRT
 <213> YiaR-Ec

<400> 30
 Met Arg Lys Ser Thr Leu Ser Gly Glu Val Arg Val Arg Asn His Gln
 1 5 10 15

Leu Gly Ile Tyr Glu Lys Ala Leu Ala Lys Asp Leu Ser Trp Pro Glu
 20 25 30

Arg Leu Val Leu Ala Lys Ser Cys Gly Phe Asp Phe Val Glu Met Ser
 35 40 45

Val Asp Glu Thr Asp Glu Arg Leu Ser Arg Leu Asp Trp Ser Ala Ala
 50 55 60

Gln Arg Thr Ser Leu Val Ala Ala Met Ile Glu Thr Gly Val Gly Ile
 65 70 75 80

Pro Ser Met Cys Leu Ser Ala His Arg Arg Phe Pro Phe Gly Ser Arg
 85 90 95

Asp Glu Ala Val Arg Glu Arg Ala Arg Glu Ile Met Ser Lys Ala Ile
 100 105 110

Arg Leu Ala Arg Asp Leu Gly Ile Arg Thr Ile Gln Leu Ala Gly Tyr
 115 120 125

Asp Val Tyr Tyr Glu Asp His Asp Glu Gly Thr Arg Gln Arg Phe Ala
 130 135 140

Glu Gly Leu Ala Trp Ala Val Glu Gln Ala Ala Ser Gln Val Met
 145 150 155 160

Leu Ala Val Glu Ile Met Asp Thr Ala Phe Met Asn Ser Ile Ser Lys
 165 170 175

Trp Lys Lys Trp Asp Glu Met Leu Ala Ser Pro Trp Phe Thr Val Tyr
 180 185 190

Pro Asp Val Gly Asn Leu Ser Ala Trp Gly Asn Asp Val Pro Ala Glu
 195 200 205

Leu Lys Leu Gly Ile Asp Arg Ile Ala Ala Ile His Leu Lys Asp Thr
 210 215 220

Gln Pro Val Thr Gly Gln Ser Pro Gly Gln Phe Arg Asp Val Pro Phe
 225 230 235 240

Gly Glu Gly Cys Val Asp Phe Val Gly Ile Phe Lys Thr Leu His Lys
 245 250 255

Leu Asn Tyr Arg Gly Ser Phe Leu Ile Glu Met Trp Thr Glu Lys Ala
 260 265 270

Lys Glu Pro Val Leu Glu Ile Ile Gln Ala Arg Arg Trp Ile Glu Ala
 275 280 285

Arg Met Gln Glu Ala Gly Phe Ile Cys
 290 295

<210> 31

<211> 286

<212> PRT

<213> YiaR-Hi

<400> 31

Met Lys Lys His Lys Ile Gly Ile Tyr Glu Lys Ala Leu Pro Lys Asn			
1	5	10	15
Ile Thr Trp Gln Glu Arg Leu Ser Leu Ala Lys Ala Cys Gly Phe Glu			
20	25	30	
Phe Ile Glu Met Ser Ile Asp Glu Ser Asn Asp Arg Leu Ser Arg Leu			
35	40	45	
Asn Trp Thr Lys Ser Glu Arg Ile Ala Leu His Gln Ser Ile Ile Gln			
50	55	60	
Ser Gly Ile Thr Ile Pro Ser Met Cys Leu Ser Ala His Arg Arg Phe			
65	70	75	80
Pro Phe Gly Ser Lys Asp Lys Lys Ile Arg Gln Lys Ser Phe Glu Ile			
85	90	95	
Met Glu Lys Ala Ile Asp Leu Ser Val Asn Leu Gly Ile Arg Thr Ile			
100	105	110	
Gln Leu Ala Gly Tyr Asp Val Tyr Tyr Glu Lys Gln Asp Glu Glu Thr			
115	120	125	
Ile Lys Tyr Phe Gln Glu Gly Ile Glu Phe Ala Val Thr Leu Ala Ala			
130	135	140	
Ser Ala Gln Val Thr Leu Ala Val Glu Ile Met Asp Thr Pro Phe Met			
145	150	155	160
Ser Ser Ile Ser Arg Trp Lys Lys Trp Asp Thr Ile Ile Asn Ser Pro			
165	170	175	
Trp Phe Thr Val Tyr Pro Asp Ile Gly Asn Leu Ser Ala Trp Asn Asn			
180	185	190	
Asn Ile Glu Glu Glu Leu Thr Leu Gly Ile Asp Lys Ile Ser Ala Ile			
195	200	205	
His Leu Lys Asp Thr Tyr Pro Val Thr Glu Thr Ser Lys Gly Gln Phe			
210	215	220	
Arg Asp Val Pro Phe Gly Gln Gly Cys Val Asp Phe Val His Phe Phe			
225	230	235	240
Ser Leu Leu Lys Lys Leu Asn Tyr Arg Gly Ala Phe Leu Ile Glu Met			
245	250	255	

Trp Thr Glu Lys Asn Glu Glu Pro Leu Leu Glu Ile Ile Gln Ala Arg
 260 265 270

Lys Trp Ile Val Gln Gln Met Glu Lys Ala Gly Leu Leu Cys
 275 280 285

<210> 32

<211> 231

<212> PRT

<213> YiaS-Ec

<400> 32

Met Leu Glu Gln Leu Lys Ala Asp Val Leu Ala Ala Asn Leu Ala Leu
 1 5 10 15

Pro Ala His His Leu Val Thr Phe Thr Trp Gly Asn Val Ser Ala Val
 20 25 30

Asp Glu Thr Arg Gln Trp Met Val Ile Lys Pro Ser Gly Val Glu Tyr
 35 40 45

Asp Val Met Thr Ala Asp Asp Met Val Val Val Glu Ile Ala Ser Gly
 50 55 60

Lys Val Val Glu Gly Ser Lys Lys Pro Ser Ser Asp Thr Pro Thr His
 65 70 75 80

Leu Ala Leu Tyr Arg Arg Tyr Ala Glu Ile Gly Gly Ile Val His Thr
 85 90 95

His Ser Arg His Ala Thr Ile Trp Ser Gln Ala Gly Leu Asp Leu Pro
 100 105 110

Ala Trp Gly Thr Thr His Ala Asp Tyr Phe Tyr Gly Ala Ile Pro Cys
 115 120 125

Thr Arg Gln Met Thr Ala Glu Glu Ile Asn Gly Glu Tyr Glu Tyr Gln
 130 135 140

Thr Gly Glu Val Ile Ile Glu Thr Phe Glu Glu Arg Gly Arg Ser Pro
 145 150 155 160

Ala Gln Ile Pro Ala Val Leu Val His Ser His Gly Pro Phe Ala Trp
 165 170 175

Gly Lys Asn Ala Ala Asp Ala Val His Asn Ala Val Val Leu Glu Glu

180

185

190

Cys Ala Tyr Met Gly Leu Phe Ser Arg Gln Leu Ala Pro Gln Leu Pro
 195 200 205

Ala Met Gln Asn Glu Leu Leu Asp Lys His Tyr Leu Arg Lys His Gly
 210 215 220

Ala Asn Ala Tyr Tyr Gly Gln
 225 230

<210> 33

<211> 231

<212> PRT

<213> YiaS-Hi

<400> 33

Met Leu Ala Gln Leu Lys Lys Glu Val Phe Glu Ala Asn Leu Ala Leu
 1 5 10 15

Pro Lys His His Leu Val Thr Phe Thr Trp Gly Asn Val Ser Ala Ile
 20 25 30

Asp Arg Glu Lys Asn Leu Val Val Ile Lys Pro Ser Gly Val Asp Tyr
 35 40 45

Asp Val Met Thr Glu Asn Asp Met Val Val Val Asp Leu Phe Thr Gly
 50 55 60

Asn Ile Val Glu Gly Asn Lys Lys Pro Ser Ser Asp Thr Pro Thr His
 65 70 75 80

Leu Glu Leu Tyr Arg Gln Phe Pro His Ile Gly Gly Ile Val His Thr
 85 90 95

His Ser Arg His Ala Thr Ile Trp Ala Gln Ala Gly Leu Asp Ile Ile
 100 105 110

Glu Val Gly Thr Thr His Gly Asp Tyr Phe Tyr Gly Thr Ile Pro Cys
 115 120 125

Thr Arg Gln Met Thr Thr Lys Glu Ile Lys Gly Asn Tyr Glu Leu Glu
 130 135 140

Thr Gly Lys Val Ile Val Glu Thr Phe Leu Ser Arg Gly Ile Glu Pro
 145 150 155 160

Asp Asn Ile Pro Ala Val Leu Val His Ser His Gly Pro Phe Ala Trp
165 170 175

Gly Lys Asp Ala Asn Asn Ala Val His Asn Ala Val Val Leu Glu Glu
180 185 190

Val Ala Tyr Met Asn Leu Phe Ser Gln Gln Leu Asn Pro Tyr Leu Ser
195 200 205

Pro Met Gln Lys Asp Leu Leu Asp Lys His Tyr Leu Arg Lys His Gly
210 215 220

Gln Asn Ala Tyr Tyr Gly Gln
225 230

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/23862

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12Q 1/68; C12P 21/06; C12N 15/00; C07H 21/02, 21/04; A61K 39/108

US CL :Please See Extra Sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 69.1, 172.3; 424/259.1, 93.1; 536/23.1, 23.4, 23.5; 935/19

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, EMBASE, BIOSIS, WEST 1.2, N. A DATA BASE SEARCH, US PATFULL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---P	US 5,824,485 A (THOMPSON et al) 20 October 1998, see the entire document.	1-9, 16, 18-20, 46-56, 63-66, 72- 74, 80, 86-87, 93, 95, 98
Y		----- 39-41, 43, 58-61, 82-85
Y	BLATTER et al. The Complete Genome Sequence of Escherichia coli K-12. 05 September 1997. vol. 277, pages 1453-1468.	39-41, 43, 58-61, 82-85

 Further documents are listed in the continuation of Box C. See patent family annex.

- * Special categories of cited documents:
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "E" document member of the same patent family

Date of the actual completion of the international search

03 FEBRUARY 2000

Date of mailing of the international search report

28 FEB 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

P. PONNAIURU

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/23862

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/23862

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/6, 69.1, 172.3; 424/259.1, 93.1; 536/23.1, 23.4, 23.5; 935/19

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-25 and 63-71, drawn to a method for screening for one or more nucleic acid sequences that express one or more products that convert a source compound into a target compound.

Group II, claim(s) 26-30, 35-38, drawn to an isolated enriched, or purified nucleic acid molecule encoding one or more yia operon-related polypeptides.

Group III, claim(s) 31-34, drawn to a nucleic acid probe for the detection of nucleic acid encoding one or more Yia operon-related polypeptides.

Group IV, claim(s) 39-45, drawn to an isolated, enriched or purified nucleic acid molecule.

Group V, claim(s) 46-62, drawn to a recombinant cell useful for screening for one or more nucleic acid sequences that express one or more products that convert a source compound into a target compound.

Group VI, claim(s) 72-81, drawn to a method of detecting the presence, absence, or amount of compound in a sample.

Group VII, claim(s) 82-85, drawn to an isolated, purified, or enriched nucleic acid molecule encoding YiaJ and a reporter gene.

Group VIII, claim(s) 86-94, drawn to a recombinant cell for detecting the presence, absence, or amount of compound in a sample.

Group IX, claim(s) 95-98, drawn to a method for selection for a nucleic acid sequence encoding a metabolic pathway from a source compound to a target compound.

The inventions listed as Groups I to IX do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of group I, nucleic acid sequences that express one or more products that convert a source compound into a target compound, is not present in the other groups; and the special technical features of other group inventions is not present in group I inventions, thus unity of invention is lacking.

THIS PAGE BLANK (USPTO)